MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 113 No. 1 1953

Published and Sold by the
ROYAL ASTRONOMICAL SOCIETY
BURLINGTON HOUSE
LONDON, W. 1

Price Thirteen Shillings and Sixpence

NOTICE TO AUTHORS

1. *Communications.*—Papers must be communicated to the Society by a Fellow. They should be accompanied by a summary at the *beginning* of the paper conveying briefly the content of the paper, and drawing attention to important new information and to the main conclusions. The summary should be intelligible in itself, without reference to the paper, to a reader with some knowledge of the subject; it should not normally exceed 200 words in length. **Authors are requested to submit MSS. in duplicate. These should be typed using double spacing and leaving a margin of not less than one inch on the left-hand side. Corrections to the MSS. should be made in the text and not in the margin.** Unless a paper reaches the Secretaries more than seven days before a Council meeting it will not normally be considered at that meeting. By Council decision, MSS. of accepted papers are retained by the Society for one year after publication; unless their return is then requested by the author, they are destroyed.

2. *Presentation.*—Authors are allowed considerable latitude, but they are requested to follow the general style and arrangement of *Monthly Notices*. References to literature should be given in the standard form, including a date, for printing either as footnotes or in a numbered list at the end of the paper. Each reference should give the **name and initials** of the author cited, irrespective of the occurrence of the name in the text (some latitude being permissible, however, in the case of an author referring to his own work). The following examples indicate the style of reference appropriate for a paper and a book, respectively :—

A. Corlin, *Zs. f. Astrophys.*, **15**, 239, 1938.
H. Jeffreys, *Theory of Probability*, 2nd edn., section 5.45, p. 258, Oxford, 1948.

3. *Notation.*—For technical astronomical terms, authors should conform closely to the recommendations of Commission 3 of the International Astronomical Union (*Trans. I.A.U.*, Vol. VI, p. 345, 1938). Council has decided to adopt the I.A.U. 3-letter abbreviations for constellations where contraction is desirable (Vol. IV, p. 221, 1932). In general matters, authors should follow the recommendations in *Symbols, Signs and Abbreviations* (London : Royal Society, 1951) except where these conflict with I.A.U. practice.

4. *Diagrams.*—**These should be designed to appear upright on the page, drawn about twice the size required in print and prepared for direct photographic reproduction except for the lettering, which should be inserted in pencil. Legends should be given in the manuscript indicating where in the text the figure should appear.** Blocks are retained by the Society for 10 years; unless the author requires them before the end of this period they are then destroyed.

5. *Tables.*—**These should be arranged so that they can be printed upright on the page.**

6. *Proofs.*—Costs of alterations exceeding 5 per cent of composition must be borne by the author. Fellows are warned that such costs have risen sharply in recent years, and it is in their own and the Society's interests to seek the maximum conciseness and simplification of symbols and equations consistent with clarity.

7. *Revised Manuscripts.*—When papers are submitted in revised form it is especially requested that they be accompanied by the original MS.

Reading of Papers at Meetings

8. When submitting papers authors are requested to indicate whether they will be willing and able to read the paper at the next or some subsequent meeting, and approximately how long they would like to be allotted for speaking.

9. Postcards giving the programme of each meeting are issued some days before the meeting concerned. Fellows wishing to receive such cards whether for Ordinary Meetings or for the Geophysical Discussions or both should notify the Assistant Secretary.

MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 113 No. 1

MEETING OF 1953 JANUARY 9

Professor H. Dingle, President, in the Chair

The election by the Council of the following Fellows was duly confirmed:—

*Alan Henry Batten, St. Regulus Hall, St. Andrews, Fife (proposed by E. Finlay Freundlich);

*Joyce Gertrude Billings, B.Sc., 43 Rockland Road, Downend, Bristol, Somerset (proposed by W. H. McCrea);

Eva Cassirer, Department of History and Philosophy of Science, University College, London, W.C.1 (proposed by H. Dingle);

Leonard William Clarke, M.A., 115 Swakeley's Road, Ickenham, Middlesex (proposed by T. G. Cowling);

*Donald George Ewart, B.Sc., 15 Lilburn Gardens, South Gosforth, Newcastle-on-Tyne, 3 (proposed by W. M. Smart);

Randolph D. Gifford, D.Sc., A.M.I.E.E., 15 Willington Road, Birmingham, 20 (proposed by M. C. Johnson);

Leonard John Heath, B.Sc., 20 Wren Gardens, Dagenham, Essex (proposed by C. A. Ronan);

Samuel Herrick, B.A., Ph.D., Department of Astronomy, University of California, Los Angeles, California, U.S.A. (proposed by D. H. Sadler);

Jan Hospers, Pembroke College, Cambridge (proposed by S. K. Runcorn);

*Angela Margaret James, H.M. Nautical Almanac Office, Herstmonceux Castle, near Hailsham, Sussex (proposed by D. H. Sadler);

*Leon Mestel, Department of Mathematics, The University, Leeds (proposed by T. G. Cowling);

Jack Miller, 10 Bury Street, St. James's, London, S.W.1 (proposed by M. Davidson);

*Peter Coles Price, B.A., Trinity College, Cambridge (proposed by R. O. Redman);

Michael John Seaton, Ph.D., 26 Kilmorie Road, London, S.E.23 (proposed by H. S. W. Massey); and

*Robert Wilson, B.Sc., Royal Observatory, Edinburgh (proposed by W. M. H. Greaves).

* Transferred from Junior Membership.

The election by the Council of the following Junior Members was duly confirmed:—

Gilbert Elliott Satterthwaite, 48 Upper Abbey Road, Belvedere, Kent (proposed by P. J. D. Gething); and

Paul Anthony Young, 125 Saltram Crescent, London, W.9 (proposed by R. W. B. Pearse).

Ninety presents were announced as having been received since the last meeting.

The President announced that the Council had awarded the Gold Medal of the Society to Professor Subrahmanyan Chandrasekhar, of Yerkes Observatory, Williams Bay, Wisconsin, U.S.A., for his contributions to mathematical astrophysics.

The President announced that the Council had awarded the Eddington Medal to Canon Georges Lemaître, of the University of Louvain, for his work on the expansion of the universe.

The President announced that the Council had awarded a Jackson-Gwilt Medal and Gift to Mr John Philip Manning Prentice for his contributions to the study of meteors.

AN INVESTIGATION OF THE H II REGIONS BY A RADIO METHOD

P. A. G. Scheuer and M. Ryle

(Received 1952 November 17*)

Summary

Observations have been made to determine the effect of the H II regions on the radio emission from the galaxy. Since both interstellar hydrogen and early-type stars show strong galactic concentration, regions of ionized hydrogen are likely to produce the greatest effect in directions near the galactic equator. At low frequencies, where the " brightness temperature " of the general galactic radiation exceeds the kinetic temperature of the interstellar hydrogen, the absorption due to the hydrogen will exceed its emission, and a dark band is to be expected along the galactic equator. At frequencies greater than about 60 Mc/s, where the brightness temperature of the galactic radiation falls below the gas temperature, this band should appear in emission.

An interferometric method of large resolving power has been used to determine the distribution of brightness near the galactic plane at frequencies of 81·5 Mc/s and 210 Mc/s. In each case a bright band about 2° wide has been observed superimposed on the general radiation. No accurate observations have yet been made at frequencies less than 60 Mc/s.

A comparison of the observed brightness temperature of the band at the two frequencies makes it possible to derive a preliminary estimate of the kinetic temperature and effective density in the H II regions. The temperature derived in this way is considerably greater than the accepted value.

Observations at different galactic longitudes have suggested that the distribution of the H II regions shows a marked concentration towards the centre of the galaxy.

---

1. *Introduction.*—The earliest theories of the radio-frequency radiation from the galaxy were based on emission from the interstellar gas (**1, 2**). Later observations at lower frequencies showed that it would be necessary to assume a mean interstellar temperature and density much greater than the values derived from the visual observations if the results obtained at frequencies between 30 and 100 Mc/s were to be explained on this hypothesis (**3**). The discovery of radio stars suggested an alternative explanation of the galactic radiation (**4, 5**) and later work has been based on the assumption that both mechanisms may contribute to the total radiation.

If the galactic radiation were due solely to radio stars, and the effects of the interstellar matter were negligible, there would be no reabsorption and the galaxy would show the characteristics of a region which was optically thin. The spectrum of the radiation would then be the same in all directions, unless there were populations of radio stars having different spectra and different distributions within the galaxy.

If the optical depth of the interstellar matter were appreciable, two effects would modify this result; firstly, the radiation from distant radio stars would suffer absorption, and secondly, the emission from the interstellar matter might

* Received in original form 1952 July 29.

make an appreciable contribution to the total radiation. If it were assumed that there were only one population of radio stars, it would therefore be possible to detect any effects of interstellar matter by observing the spectrum of the galactic radiation in a number of different directions.

Early observations by Moxon (**6**) indicated that between frequencies (ν) of 45 and 200 Mc/s, the brightness temperature of the radiation from near the galactic poles varied as $\nu^{-2\cdot1}$, whilst near the galactic centre the brightness temperature varied as $\nu^{-2\cdot7}$. These results, though they were of great interest, were not conclusive, owing to possible errors involved in making such observations with aerial systems of low resolving power.

In a more recent paper Piddington (**7**) has summarized the results of a number of measurements which have been made to determine the distribution of intensity with direction at frequencies ranging from 18 to 3000 Mc/s. According to his results the contours of constant intensity near the galactic plane show a progressive broadening at the lower frequencies; if the derived contours at 64 and 18 Mc/s are correct it would be necessary to postulate that the optical depth of the interstellar gas was appreciable even in directions 15° away from the galactic equator.*

Westerhout and Oort (**9**), on the other hand, have concluded that any interstellar emission and absorption processes must be associated with free-free transitions in regions of ionized hydrogen in the neighbourhood of early-type stars; in all other interstellar regions either the temperature or the density is too low for appreciable radio-frequency emission. These ionized hydrogen regions have been observed by their Balmer-line emission spectra (**10**) within a few kiloparsecs of the Sun, and the observations have shown that most of the regions are within about 200 parsecs of the galactic plane. If radio observations are made from the Earth, the total optical depth of the galaxy due to such regions should therefore decrease rapidly with increasing galactic latitude; on present evidence the optical depth might be expected to fall to half at about 1° from the galactic equator. Because of its limited extent Westerhout and Oort concluded that the interstellar matter would be unlikely to produce an effect sufficient to be detected with the comparatively poor resolving power which was used in the surveys they considered.

The present paper describes an investigation which has been made to detect the presence of a distribution of regions of ionized hydrogen similar to that considered by Westerhout and Oort. An interferometric method of large resolving power has been used to obtain detailed measurements of the intensity distribution near the galactic plane, at frequencies of 81·5 and 210 Mc/s; the results of these measurements have then been compared with the predicted distribution of intensity, taking into account the general galactic radiation, and the emission and absorption associated with the regions of ionized hydrogen.

2. *Summary of theoretical analysis.*—The intensity of radio waves from a given direction depends on the emission from radio stars, the interstellar gas, and extragalactic sources of radiation, and on the absorption of all components of the radiation in the interstellar gas. Westerhout and Oort have shown that the main part of the galactic radiation is consistent with emission from radio stars, distributed according to the model of Oort and van Woerkom (**11**). It is, however, possible that "radio stars" are in fact situated outside the galaxy, and that some other

* Preliminary observations (**8**) at frequencies of 72 and 158 Mc/s, using the larger resolving power of the 220-ft. paraboloid at Jodrell Bank, have indicated that there is little difference between the contours at these two frequencies in a region between galactic longitudes 60° and 120°.

mechanism is responsible for the radiation from the galaxy. In that case, the analysis of this section would still be applicable, provided that the new mechanism could be regarded as being "optically thin". From considerations such as those of Ryle (**12**) and Hutchinson(**13**) it seems likely that any interstellar mechanism for galactic radiation, other than that of free-free transitions, would be associated with particles of very great energy, and would therefore satisfy the requirement of small optical depth. For the present purposes it will be convenient to assume that radio stars are in fact responsible for the main component.

In order to explain the intensity distribution observed by Bolton and Westfold (**14**) at 100 Mc/s, Westerhout and Oort further postulated the existence of an isotropic component of radiation, having an intensity about 10 per cent of that observed in the direction of the galactic centre. It is not yet certain whether this component represents a true extragalactic contribution, or whether the observations could be accounted for by a more spherical distribution of radio stars; some evidence for the former hypothesis has recently been obtained by Hanbury Brown and Hazard (**15**).

The effect of the H II regions on the intensity distribution will now be considered. The absorption coefficient for a frequency ν in a region containing N_e electrons, and N_i ions of atomic number Z, per cm^3, is given by Smerd and Westfold (**16**):

$$\kappa_\nu = \frac{4e^6 A_1(2) N_e N_i Z^2}{3(2\pi)^{1/2}(m_e k T_e)^{3/2} C \nu^2 \mu_\nu},$$

where μ_ν is the refractive index (which may be taken as 1 in all the regions discussed here), T_e is the kinetic temperature in the region, and

$$A_1(2) = \ln \{1 + (4kT_e/e^2 N_e^{1/3})^2\}.$$

Following Westerhout and Oort, $Z^2 N_i/N_e$ is taken to be 1·1. For $T_e = 10^4$ deg. K and $N_e = 8$, $A_1(2) = 32{\cdot}6$. If N_e is changed by a factor of 10 or T_e by a factor of 2, $A_1(2)$ is only changed by about 5 per cent, and $A_1(2)$ is therefore assumed to be 32·6 throughout the following calculations.

The absorption coefficient can now be written

$$\kappa_\nu = 0{\cdot}177\, T_e^{-3/2}\, \nu^{-2} N_e^2.$$

It can be shown that, when the optical depth of a single H II cloud is small, it is permissible to treat a region containing many clouds as a continuous medium with an effective absorption coefficient given by substituting, in the above formula, the mean value of N_e^2 over the clouds and the spaces between them. Westerhout and Oort estimate that the optical depth of a single cloud is 0·02 at 100 Mc/s, so that at 30 Mc/s it would be about 0·2; this approximation is therefore certainly valid for frequencies greater than 30 Mc/s.

Now suppose that, in a particular direction, the brightness temperature due to the radio stars in the absence of absorption is T_s, and that the contribution to this temperature from radio stars situated at distances between s and $s+ds$ is $t_s ds$. Suppose, also, that the extragalactic component produces a brightness temperature T_s'. If the kinetic temperature (T_e) is assumed to be uniform throughout all the H II regions, the actual brightness temperature T is given by

$$T = T_e \left\{1 - \exp\left(-\int_0^\infty \kappa_\nu\, ds\right)\right\} + \int_0^\infty t_s \exp\left(-\int_0^s \kappa_\nu\, ds_1\right) ds + T_s' \exp\left(-\int_0^\infty \kappa_\nu\, ds\right).$$

In the absence of absorption and emission by the H II regions, the brightness temperature would be simply $T^* = T_s + T_s'$. The presence of the H II regions thus increases the observed brightness temperature by an amount

$$(T - T^*) = T_e \left\{1 - \exp\left(-\int_0^\infty \kappa_\nu \, ds\right)\right\} - \int_0^\infty t_s \left\{1 - \exp\left(-\int_0^s \kappa_\nu \, ds_1\right)\right\} ds - T_s' \left\{1 - \exp\left(-\int_0^\infty \kappa_\nu \, ds\right)\right\}. \quad (1)$$

The first and third terms of the right-hand member of equation (1) depend only on $\tau_\nu = \int_0^\infty \kappa_\nu \, ds$, the total optical depth of the H II regions in the direction considered, but the second term (the absorption of radio-star radiation) also depends to some extent on the distribution of the H II clouds and the radio stars along the line of sight. For calculating this term, it will be assumed initially that the H II regions are distributed uniformly over a disk of radius 12·5 kps, and that the radio stars are distributed according to a simplified version of Oort and van Woerkom's model, due to de Kort (**17**).*

Equation (1) can now be expressed in the form

$$(T - T^*) = T_e \{1 - \exp(-\tau_\nu)\} - f(\tau_\nu) T^*, \quad (2)$$

where $f(\tau_\nu)$ is a known function of τ_ν for a given galactic longitude.

Using the constants adopted by Strömgren (**19**),

$$T_e = 10^4 \text{ deg. K}, \quad \int N_e^2 \, ds = 1{\cdot}9 \times 10^{22} \text{ cm}^{-5},$$

the emission from the H II regions (first term of equation (2)) and the absorption of the radio-star and extragalactic components (second term of equation (2)) have been computed for a direction near the galactic centre. These curves, together with their sum, which represents the overall effect of the H II regions, $(T - T^*)$, are shown in Fig. 1. It can be seen that with these constants, the H II regions should be visible in emission at frequencies greater than 60 Mc/s, and in absorption at lower frequencies. At other galactic longitudes the transition between emission and absorption should take place at a lower frequency.

If measurements of the brightness temperature at a given galactic longitude were made at different galactic latitudes, a curve like Fig. 2 would be obtained, from which both T^* and $(T - T^*)$ could be estimated. Using equation (2), a series of theoretical curves could be constructed relating the value of $(T - T^*)$ to T_e and $\int N_e^2 \, ds$. For a different frequency or a different value of T^* another set of curves would be obtained : examples of such curves (for $T^* = 6\,500$ deg. K at 81·5 Mc/s and $T^* = 750$ deg. K at 210 Mc/s) are shown in Fig. 3. From the point of intersection of the curves representing the observed values of $(T - T^*)$ at two different frequencies, it would then be possible to determine both T_e and $\int N_e^2 \, ds$ for the direction of observation.

If, now, such observations were made at a number of galactic longitudes, it would become possible to obtain information on the actual spatial distribution of the H II regions; it would then no longer be necessary to assume the model of a uniform disk.

* In a recent paper Mills (**18**) concludes that there is a significant concentration of intense radio stars towards the galactic equator, and suggests that these may represent a new class of radio stars showing strong galactic concentration. The significance of these observations in relation to the present investigation will be discussed in Section 7.

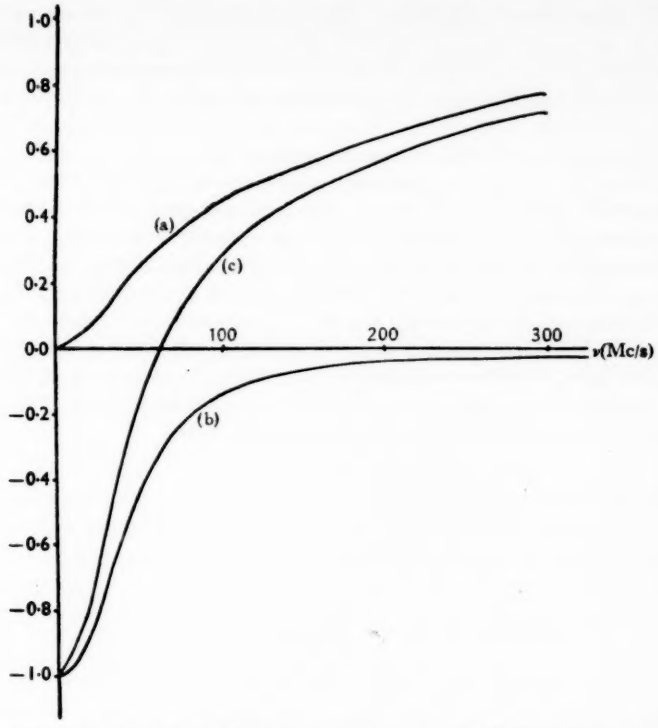

FIG. 1.—*Estimated changes in brightness temperature due to H II regions, expressed as fractions of T*, the brightness temperature which would be observed in the absence of H II regions.*

(a) Emission from H II regions.
(b) Absorption of "radio-star" and "extragalactic" components.
(c) Net change due to H II regions.

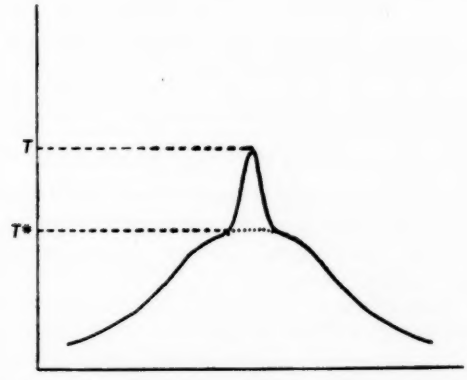

FIG. 2.—*Sketch of the expected variation of brightness temperature with galactic latitude.*
Ordinates: Brightness temperature.
Abscissae: Galactic latitude (b).

3. *The experimental difficulties.*—Before describing the present method of observation, it may be useful to indicate why previous surveys of galactic radiation have failed to disclose the existence of a feature 2° wide, and probably more than 90° long, having a brightness temperature perhaps twice as great as that of neighbouring areas of the sky. The two most complete surveys of galactic radiation (**14, 20**) were made with aerial systems having apertures of the order of four wave-lengths, giving reception patterns whose angular widths were about 15°. The effect of the finite width of the reception pattern has been carefully corrected in these surveys, but it is possible to show that such methods cannot yield information on angular structure smaller than about $1/2n$ radians, where $n\lambda$ is the effective aperture of the aerial. Only two surveys appear to have been made with a sufficiently large resolving power: one of them, reported by Lovell and Clegg (**8**), made use of the 220-ft. paraboloid at Jodrell Bank, but was restricted to areas within 15° of the zenith, and therefore gave no information on regions near the galactic centre; the other was made by Piddington and Minnett (**21**), at a frequency of 1 200 Mc/s, in a restricted region near the galactic centre.

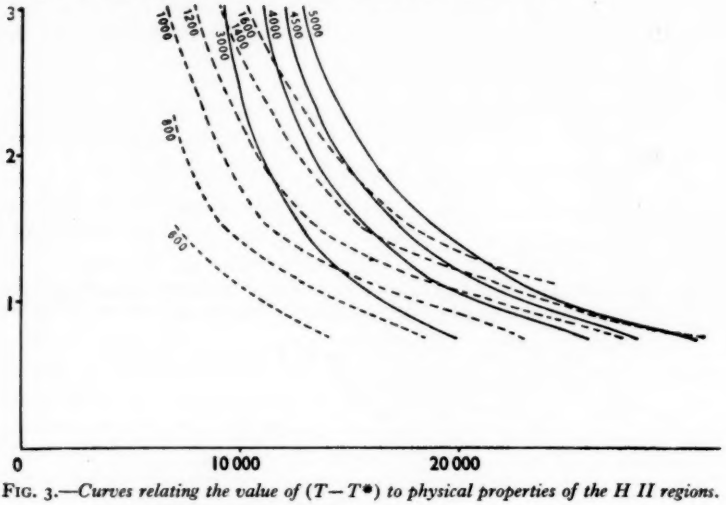

FIG. 3.—*Curves relating the value of* $(T-T^{\bullet})$ *to physical properties of the H II regions.*
Ordinates: $T_e^{-3/2} \int N_e^2\, ds \times 10^{-6}$.
Abscissae: Kinetic temperature of hydrogen clouds (T_e).
Curves for 81·5 *Mc/s:* ————
Curves for 210 *Mc/s:* — — — —
Each curve is labelled with the corresponding value of $(T-T^{\bullet})$.

It can be seen from Fig. 1 that an investigation of the H II regions is likely to involve accurate measurements of the intensity distribution at wave-lengths of several metres; at these wave-lengths there are serious practical difficulties in constructing aerials of sufficient aperture.

The possibility of determining the distribution of brightness across a source by observations with an interferometer of variable spacing was first pointed out by Michelson and Pease (**22**), and the application of this principle to radio astronomy was suggested by McCready, Pawsey and Payne-Scott (**23**). The method has been used by Stanier (**24**), Machin (**25**), and O'Brien (**26**) for determining the distribution of brightness across the undisturbed solar disk.

In the present problem a large resolving power is needed to determine the distribution of brightness temperature in galactic latitude, but the expected variation with galactic longitude is comparatively small; the observations could therefore be made with an aerial system having a relatively small resolving power in directions parallel to the galactic plane.

The principle of the method which has been used will be discussed in the next section; the practical details are described in Section 5.

4. *The variable spacing interferometer.*—Suppose that two aerials are mounted with a spacing of n wave-lengths, on an axis which passes through the galactic poles, and that they are directed at a point on the galactic equator ($b=0$) whose galactic longitude is l_0. Let the reception pattern of each aerial be given by $f(l, b)$.

If the distribution of brightness is represented by the brightness temperature $T(l, b)$, the power intercepted by each aerial from an element of solid angle $\delta\omega$ at (l, b) may be represented by an effective aerial temperature

$$\delta T_A = \frac{T(l,b)\,\delta\omega f(l,b)}{\iint f(l,b)\cos b\,dl\,db}.$$

The denominator of this expression may be written as $4\pi/G$, where G is the gain of each aerial over an isotropic radiator. If the two aerials are now used in a phase-switching interferometer system (**27**), the deflection on the record due to the small element is given by

$$C\delta T_A = C\frac{G}{4\pi}T(l,b)f(l,b)\,\delta\omega\cos(2\pi n\sin b),$$

where C is a constant depending on the deflection sensitivity of the recording equipment.

The deflection produced by the radiation from the whole sky may then be written

$$d(n) = C\frac{G}{4\pi}\int T(l,b)f(l,b)\cos(2\pi n\sin b)\,d\omega. \tag{3}$$

If the variation of $T(l, b)$ with l can be regarded as linear within the reception pattern of the aerials, and if $f(l, b)$ can be expressed in the form $L(l)\,B(b)$, where $L(l)$ is symmetrical about $l=l_0$, equation (3) can be rewritten

$$\begin{aligned} d(n) &= C\frac{G}{4\pi}\int_{-\pi}^{\pi} L(l)\,dl\int_{-\pi/2}^{\pi/2} T(l_0,b)\,B(b)\cos(2\pi n\sin b)\cos b\,db \\ &= C\frac{G}{4\pi}\int_{-\pi}^{\pi} L(l)\,dl\int_{-\infty}^{\infty} X(x)\cos 2\pi nx\,dx, \end{aligned} \tag{4}$$

where

$$x=\sin b,$$

$$X(x)=\begin{cases} T(l_0,b)\,B(b) & |x|<1 \\ \text{or } 0 & |x|>1. \end{cases}$$

By using a number of different aerial spacings n, $d(n)$ can be determined as a function of n. If the reception pattern of each aerial and the distribution in galactic latitude of the brightness temperature are both symmetrical, $X(x)$ is an even function of x, and it follows from (4) that

$$T(l_0,b)\,B(b)=X(x)=\frac{8\pi}{CG\int_{-\pi}^{\pi}L(l)\,dl}\int_0^{\infty} d(n)\cos 2\pi nx\,dn. \tag{5}$$

Hence it is possible to derive the required distribution of brightness temperature.

When the distribution of T is required only for values of b up to a few degrees, it is sufficiently accurate to put $x = \sin b = b$ and $B(b) = 1$.

If it is not assumed that $X(x)$ is an even function, it can still be expressed as the sum of an even and an odd function; the former will be given by (5), and the latter by the corresponding Fourier sine transform. This sine transform could, in principle, be found by inserting an extra quarter wave-length of cable into one side of the interferometer; the cosines in (3), (4) and (5) then all become sines, and the right-hand side of (5) gives the antisymmetrical part of $X(x)$.

In practice, however, both the cosine and the sine transform of $X(x)$ can usually be obtained from the same record in the following way. The above calculations refer to the deflection $d(n)$ at the instant when the interferometer axis passes through the galactic poles; owing to the rotation of the Earth the aerials will, after a short time, no longer be directed to the point $(l_0, 0)$, but to $(l_0 + l_1, b_1)$, and the deflection is now

$$C\frac{G}{4\pi}\int T(l,b)f(l-l_1, b-b_1)\cos(2\pi n \sin(b-b_1))\,d\omega. \qquad (3a)$$

Provided l_1 and b_1 are small they can be neglected in all factors except the last; and furthermore, for small b, $\sin(b-b_1) = b-b_1$.

The deflection is therefore approximately

$$C\frac{G}{4\pi}\int_{-\pi}^{\pi} L(l)\,dl\left\{\left[\int_{-\pi/2}^{\pi/2} T(l_0,b)\,B(b)\cos 2\pi n b\,db\right]\cos 2\pi n b_1 + \left[\int_{-\pi/2}^{\pi/2} T(l_0,b)\,B(b)\sin 2\pi n b\,db\right]\sin 2\pi n b_1\right\}. \qquad (4a)$$

Thus except for very small values of n the variation with time is approximately sinusoidal, and the amplitude and phase of the record enable both Fourier transforms of $T(l_0, b)\,B(b)$ to be found.

5. *The experimental method.* (*a*) *The aerial system.*—In the method described above, the resolving power in galactic latitude does not depend on the size of the aerials, but simply on the greatest spacing used. In order to obtain information on the distribution in galactic longitude, the aerials were made with the greatest practicable aperture in a direction parallel to the galactic plane. As the aerials must be moved many times during the experiments this dimension was limited to about 9 m. At 210 Mc/s the reception pattern in galactic longitude covered about 12°; at 81·5 Mc/s it was about 30°. The largest spacing between the aerials was about 60 wave-lengths in both experiments. Additional observations on 81·5 Mc/s were made with a recording system capable of measuring the total aerial power (**28**), so that the zero-order Fourier component could be determined. No measurements of total aerial power were made on 210 Mc/s, and the results therefore only give the variation of intensity over a few degrees on either side of the galactic plane.

Owing to the large spacing employed (up to 200 metres), it is important that the line joining the aerials shall be horizontal. Observations were therefore made at times when the galactic plane was vertical; that is, when one of the galactic poles was rising and the other setting. This situation occurs twice daily at all places where the galactic poles are not circumpolar; at Cambridge, the axis joining the

galactic poles is 40° west or east of north at these times.* The former of these axes was used as the interferometer axis, since observations can then be made nearer to the galactic centre; unfortunately the galactic centre itself (at galactic longitude 327°) cannot be observed at Cambridge by this method.

By using the aerials at various angles of elevation, the average distribution in galactic latitude was found for regions centred on galactic longitudes between 338° and 16°. A few observations were also made near the galactic anticentre. Observations at intermediate angles were prevented by the confusion due to the intense radio stars in Cygnus and Cassiopeia.

In the present preliminary investigations no attempt was made to determine the phase of the recorded pattern and it was assumed that the distribution was symmetrical about some plane near the assumed galactic plane. More refined experiments should prove valuable not only in investigating any asymmetry in the distribution of brightness temperature, but also in providing a method of determining the position of the galactic poles from observations of objects showing high galactic concentration and situated in the most distant parts of the galaxy.

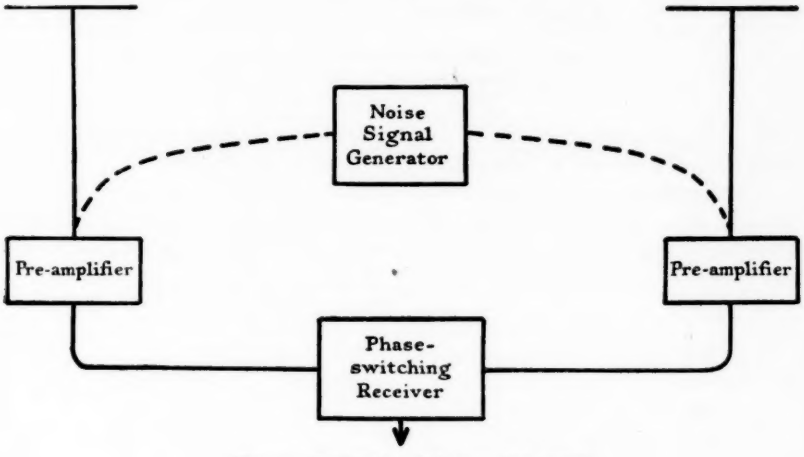

FIG. 4.—*Calibration of the recording system.*

(*b*) *Calibration of the recording system.*—A full account of the phase-switching system, including the use of pre-amplifiers at the aerials, has been given elsewhere (**27**). The deflection sensitivity C was found by connecting the output of a noise signal generator to the pre-amplifiers through the two equal aerial cables; the output of the signal generator then corresponds to the aerial power produced by a source of small angular diameter situated in the central maximum of the interference pattern.

At 81·5 Mc/s, the total aerial power is comparable with that generated in the pre-amplifiers, and if, as is usual with the phase-switching system, the gain of the receiver is controlled automatically by the total output power from the pre-amplifiers, the gain will not be the same during the calibration. The small error due to this effect was measured and corrected in the manner explained by Ryle (**27**).

* Assuming the galactic poles to be at declination ±28°.

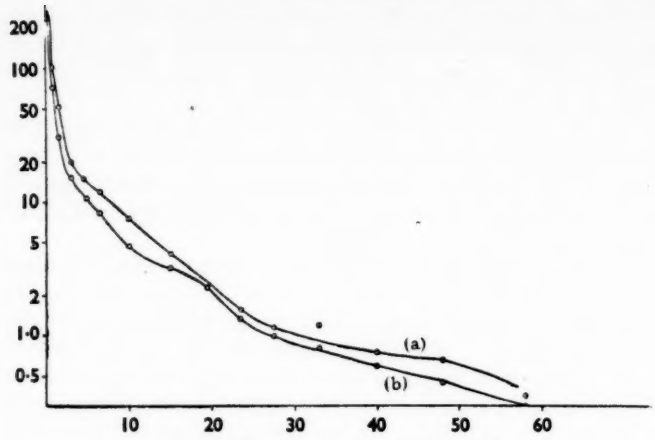

FIG. 5.—*Amplitude–spacing curves at 81·5 Mc/s.*
(*a*) *Galactic longitude* $l=353°$.
(*b*) $l=13°$.
Ordinates: Amplitudes of record (arbitrary units).
Abscissae: Spacing between aerials, n (wave-lengths).

6. *The experimental results.*—The graphs of amplitude against spacing shown in Figs. 5 and 6 contain the immediate results of the experiment.

The distributions of brightness temperature deduced from these curves are shown in Figs. 7 and 8. It is considered unlikely that the overall calibration of the recording system and the calculation of the aerial reception pattern could introduce systematic errors greater than 20 per cent.

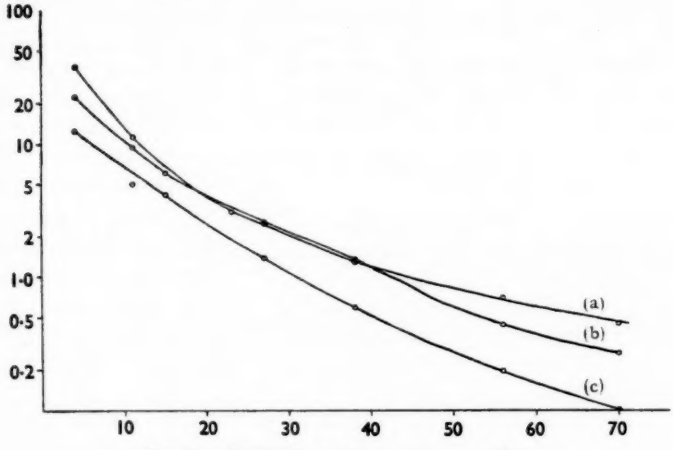

FIG. 6.—*Amplitude–spacing curves at 210 Mc/s.*
(*a*) $l=338°$.
(*b*) $l=358°$.
(*c*) $l=16°$.
Ordinates: Amplitude of record (arbitrary units).
Abscissae: Spacing between aerials, n (wave-lengths).

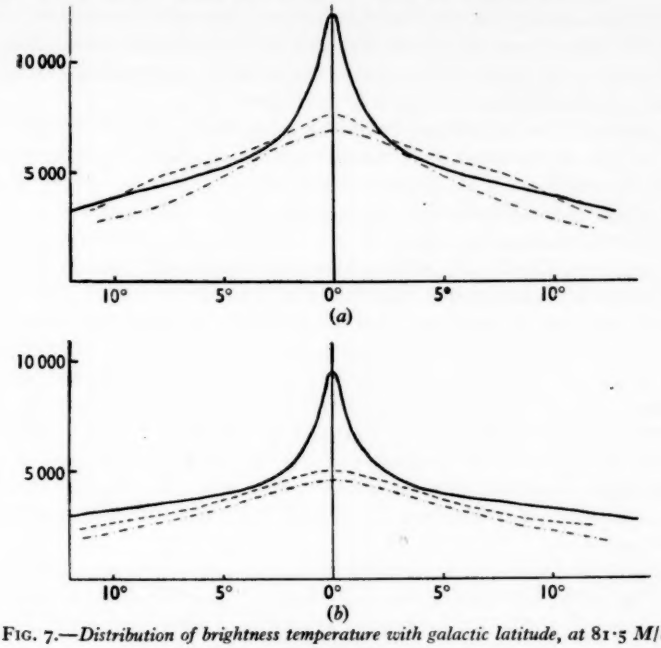

FIG. 7.—*Distribution of brightness temperature with galactic latitude, at 81·5 M/cs.*

(a) Galactic longitude $l=353°$.
(b) $l=13°$.
Ordinates: Brightness temperature (deg. K).
Abscissae: Galactic latitude.
Experimental curves: ————
Bolton and Westfold survey: — — — — —
Westerhout and Oort: —·—·—·

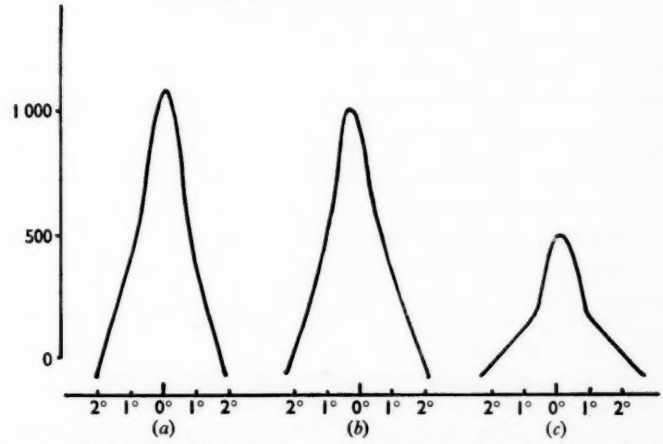

FIG. 8.—*Distribution of brightness temperature with galactic latitude at 210 Mc/s.*

(a) $l=338°$.
(b) $l=358°$.
(c) $l=16°$.
Ordinates: Brightness temperature scale (deg. K).

For comparison with the present observations, the corresponding curves from the survey of Bolton and Westfold (**14**) and the theoretical curves derived by Westerhout and Oort (**9**) for the contribution of radio stars and the extragalactic component are also shown in Fig. 7.

The values of T^* were obtained from Fig. 7 by drawing a smooth curve having the same form as that derived by Westerhout and Oort, to extrapolate the experimental curves outside $b = 3°$ to the galactic equator.

At 210 Mc/s no measurements of total power were made, and no observations were made with spacings less than four wave-lengths; the distributions deduced are therefore only reliable within a few degrees of the galactic equator and are only shown up to $b = 2°$. In order to derive the magnitudes of $(T - T^*)$ at 210 Mc/s from Fig. 8, the angular width of the H II regions as observed at 81·5 Mc/s was used.

Some additional observations at 210 Mc/s were made towards the anticentre; these showed that $(T - T^*)$ in that direction is probably less than 100 deg. K.

(*a*) *The angular distribution.*—A comparison of Figs. 7 and 8 with curves given by Westerhout and Oort shows that the distribution in latitude is in good agreement with that predicted theoretically; the width of the region found experimentally is, however, slightly greater than that predicted from the distribution of O- and B-type stars.

Although observations have not yet been made at a sufficient number of galactic longitudes for a detailed comparison to be made, it is interesting to compare the observed variation of $(T - T^*)$ with galactic longitude with that deduced from the two models suggested by Westerhout and Oort. The first of these models assumed that the H II regions were contained within a ring whose inner and outer radii were 5 and 11 kiloparsecs; the second assumed that they were distributed uniformly over a disk of radius 12·5 kiloparsecs.

The expected variation with galactic longitude of $(T - T^*)$ at 210 Mc/s is shown for the two models in Fig. 9 together with the observed values. The crosses show the observed values and indicate the estimated limits of error and the range of longitude covered by the aerials. It can be seen that the observed distribution shows an even greater central concentration than the disk model.

The absence of an appreciable component in directions near the anti-centre also provides some evidence in favour of a rather restricted radial extent.

(*b*) *The kinetic temperature and effective density of the H II regions.*—If the results are now used in the analysis described in Section 2, it is found that they lead to a kinetic temperature T_e of at least 18 000 deg. K. (This figure was derived from a high estimate (1 400 deg. K) of the 210 Mc/s temperature, and a low estimate (4 200 deg. K) of the 81·5 Mc/s temperature.) The corresponding value of $\int N_e^2\, ds$ is 3×10^{22} cm^{-5}.

These results may be compared with the values used by Westerhout and Oort in their theoretical analysis:

$$T_e = 10\,000 \text{ deg. K}, \quad \int N_e^2\, ds = 1{\cdot}9 \times 10^{22} \text{ cm}^{-5}.$$

An examination of the form of the curves shown in Fig. 3 shows that for kinetic temperatures as great as 18 000 deg. K, the present two frequencies do not offer the possibility of a very accurate determination of T_e; better accuracy could, in this case, be obtained by making observations at lower frequencies. Nevertheless it appears impossible to account for the present observations unless the kinetic temperature is considerably greater than the value previously assumed.

7. *Discussion.*—The experiments described in this paper were made in an attempt to find the radio-frequency effects of the H II regions. Such effects should, according to numerous investigations at optical wave-lengths, be confined to a narrow belt close to the galactic plane, and a narrow bright belt near the galactic plane has indeed been found in experiments at two different frequencies; the width of this belt is only slightly greater than that predicted from observations at optical wave-lengths.

There is, however, a marked difference between the distribution found in galactic longitude and that predicted theoretically. Observations of extragalactic nebulae at optical wave-lengths (**29**) have suggested that the H II regions are largely confined to the spiral arms; they have not been observed in the nuclear regions. The present observations, on the other hand, suggest that there is a marked concentration towards the centre of the galaxy.

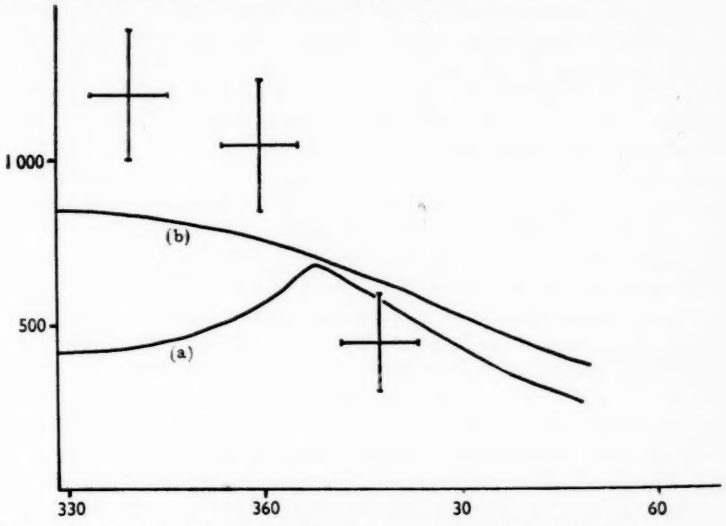

FIG. 9.—*Comparison of observed and predicted distribution of* $(T-T^*)$ *in galactic longitude.*

(a) Ring model.
(b) Disk model.
Ordinates: Brightness temperature (deg. K).
Abscissae: Galactic longitude (l).

Estimates of the average electron density of the H II regions do not differ significantly from those adopted by Westerhout and Oort, but the deduced kinetic temperature ($\lesssim$18 000 deg. K) is considerably greater; Spitzer and Savedoff (**30**) estimate that the temperature in an H II region should lie in the range 6 000–13 000 deg. K.

In view of this discrepancy, it is important to examine the possibility that some other source of radiation may be responsible for the observed bright belt along the galactic equator.

It is possible that one of the two intense radio stars in the constellations of Cassiopeia and Cygnus might produce by means of a subsidiary maximum of the reception pattern an appreciable signal in the aerial which would be interpreted as a source of restricted width; both of these sources are however situated at considerably greater declinations than the region under investigation, and the records would

therefore readily be distinguished by their longer period. All of the radio stars whose positions lie within the limits of the bright belt have intensities far too small to make a significant contribution, individually.

Recently, however, Mills (**18**) has found evidence for a concentration of intense radio stars near the galactic equator, and he has suggested that they may represent a population of "Class I" radio stars having a real concentration towards the galactic plane; it is therefore important to see whether the integrated radiation from such a population of radio stars could account for the observed bright belt. It is first of all clear that the spectrum of the radiation produced in this way would be the same as that of an individual radio star. The majority of radio stars have a spectrum which would give rise to a brightness temperature proportional to $\lambda^{2\cdot7}$ (**31**); the radio star in Taurus has been reported (**21**) to have a brightness temperature proportional to $\lambda^{2\cdot0}$, but there is, as yet, no evidence in favour of a general law of this type for those radio stars situated near the galactic equator.

The brightness temperature of the observed belt has been found to vary, between 210 Mc/s and 81·5 Mc/s, with an exponent *less* than 2·0, as would be consistent with an origin in the H II regions. Furthermore, some recent preliminary observations at 38 Mc/s have shown that between 81·5 and 38 Mc/s the excess brightness temperature *decreases* with wave-length. It does not therefore appear possible to account for the present observations in terms of the emission from a new population of radio stars, unless these are further supposed to have a spectrum very different from the known radio stars.

It is possible that the observed belt might be accounted for partly by the emission from H II regions having a temperature of the order of 10 000 deg. K, and partly by the emission from a new population of radio stars; such a model allows considerable freedom as regards spectrum, and might also allow the necessary concentration in galactic latitude. However, if Mills's value of the population density of the "Class I" radio stars is used to estimate the intensity of the integrated radiation, it is found that it could only account for a few per cent of the observed belt. Furthermore, if the estimate of density is increased, the integrated radiation is found to vary only as $\ln 1/d$, where d is the average separation between stars, whilst at the same time the degree of galactic concentration must be increased if the width of the belt is to remain sufficiently small.

Whilst a more definite conclusion should be possible when more complete information is available on the distribution and spectra of a large number of radio stars, it is clear that with the present data there are very serious difficulties in explaining the observed belt in terms of radiation from radio stars.

The most powerful evidence for an origin in the H II regions would be provided by an observation of the belt in absorption, but if the kinetic temperature is greater than 18 000 deg. K it will be necessary to make observations at frequencies considerably lower than 60 Mc/s. Some preliminary measurements at 38 Mc/s have shown that the belt is still just visible in emission; a kinetic temperature of about 25 000 deg. K would account for these observations and would also be consistent with the measurements at 81·5 Mc/s and 210 Mc/s.

Some additional supporting evidence for the contribution of the H II regions to galactic radio emission has been provided by other observations: the constants derived from the present experiments appear to account for the differences in the galactic contours obtained at 100 Mc/s (**14**) and 200 Mc/s (**20**), and would give intensities at 1 200 Mc/s and 3 000 Mc/s comparable with those found by

Piddington and Minnett (**21**). Furthermore, as pointed out by Piddington and Minnett (**32**), the extended source "Cygnus X", which lies near the direction of a spiral arm, has a spectrum similar to that expected from ionized hydrogen. Since the H II regions are found to be concentrated into spiral arms (**29**), all the available evidence suggests that the radiation from "Cygnus X" may be the integrated radiation from all the H II regions in the spiral arm.

We should like to express our thanks to our colleagues in the Laboratory for assistance in making the observations, and to Mrs Moore for help in the computation of the results.

This work was carried out as part of a programme of radio research at the Cavendish Laboratory, supported by a grant from the Department of Scientific and Industrial Research, to which one of us (P. A. G. S.) is also indebted for a Maintenance Allowance.

Cavendish Laboratory,
Cambridge :
1952 *November* 15.

References

(1) L. G. Henyey and P. C. Keenan, *Ap. J.*, **91**, 625, 1940.
(2) H. C. van de Hulst, *Ned. Tijdschr. v. Natuurkunde*, **11**, 210, 1945.
(3) C. H. Townes, *Ap. J.*, **105**, 235, 1946.
(4) A. Unsöld, *Nature, Lond.*, **163**, 489, 1949.
(5) M. Ryle, *Rep. Prog. Phys.*, **13**, 184, 1950.
(6) L. A. Moxon, *Nature, Lond.*, **158**, 758, 1946.
(7) J. H. Piddington, *M.N.*, **111**, 45, 1951.
(8) A. C. B. Lovell and J. A. Clegg, *Radio Astronomy*, p. 183, Chapman and Hall, London, 1952.
(9) G. Westerhout and J. H. Oort, *B.A.N.*, **11**, 323, 1951.
(10) O. Struve, C. T. Elvey and W. Linke, *Ap. J.*, **90**, 301, 1939.
(11) J. H. Oort and A. J. J. van Woerkom, *B.A.N.*, **9**, 185, 1941.
(12) M. Ryle, *Proc. of Con. on Dyn. of Ion. Med.*, London, 1951.
(13) G. W. Hutchinson, *Phil. Mag.*, **43**, 847, 1952.
(14) J. G. Bolton and K. C. Westfold, *Aust. J. Sci. Res.*, A**3**, 19, 1950.
(15) R. Hanbury Brown and C. Hazard, *Phil. Mag.*, **43**, 137, 1952.
(16) S. F. Smerd and K. C. Westfold, *Phil. Mag.*, **40**, 831, 1949.
(17) J. de Kort, *B.A.N.*, **9**, 189, 1941.
(18) B. Y. Mills, *Aust. J. Sci. Res.*, A**5**, 266, 1952.
(19) B. Strömgren, "*Problems of Cosmical Aerodynamics*", report of a symposium held in Paris, August, 1949.
(20) C. W. Allen and C. S. Gum, *Aust. J. Sci. Res.*, A**3**, 224, 1950.
(21) J. H. Piddington and H. C. Minnett, *Aust. J. Sci. Res.*, A**4**, 459, 1951.
(22) A. A. Michelson and F. G. Pease, *Ap. J.*, **53**, 249, 1921.
(23) L. L. McCready, J. L. Pawsey and R. Payne-Scott, *Proc. Roy. Soc.* A, **190**, 357, 1947.
(24) H. M. Stanier, *Nature, Lond.*, **165**, 354, 1950.
(25) K. E. Machin, *Nature, Lond.*, **167**, 889, 1951.
(26) P. A. O'Brien, *M.N.*, in press.
(27) M. Ryle, *Proc. Roy. Soc.* A, **211**, 351, 1952.
(28) K. E. Machin, M. Ryle and D. D. Vonberg, *Proc. I.E.E.*, **99**, Part III, 127, 1952.
(29) W. Baade and N. U. Mayall, "*Problems of Cosmical Aerodynamics*," report of a symposium held in Paris, August 1949.
(30) L. Spitzer, Jr. and M. Savedoff, *Ap. J.*, **111**, 593, 1950.
(31) G. J. Stanley and O. B. Slee, *Aust. J. Sci. Res.*, A**3**, 234, 1950.
(32) J. H. Piddington and H. C. Minnett, *Aust. J. Sci. Res.*, A**5**, 17, 1952.

CINEMATOGRAPHY OF PARTIAL SOLAR ECLIPSES

I. General

R. d'E. Atkinson

(Communicated by the Astronomer Royal)

(Received 1952 November 6)

Summary

When a central eclipse of the Sun is viewed from a station slightly outside the central track, the thin unobscured crescent changes its position angle rapidly near mid-eclipse; a detailed record of its behaviour should provide good corrections to the Moon's relative place in both coordinates. An expedition was sent to Mombasa for the eclipse of 1948 November 1, and obtained a successful timed cinema record. The present paper is an introduction to this work, and deals with general principles, choice of apparatus, choice of site, and the suitability of different eclipses. If only the P.A. of the line of cusps is measured, small images are adequate, and the camera can be stationary; if the light is reflected into it from a plane mirror, it can usually also be horizontal. Formulae are developed for the azimuth (and, if necessary, tilts) required to make the image move parallel to the long edge of the frame. The optical features of the Mombasa apparatus are outlined. Convenient approximate formulae are developed for the rate of rotation of the crescent, the sharpness of the cusps, and their divergence from each other, so that the suitability of a site can be rapidly assessed in any eclipse; the maximum admissible sharpness and divergence are discussed, and eclipse statistics are summarized from this point of view. Both total and annular eclipses are suitable.

If an observer be stationed near the central line of an eclipse of the Sun, but a little outside the actual track of totality (or of annularity), he will naturally see a thin crescent of unobscured solar surface even at mid-eclipse. It is easily verified that the position angle taken up by this crescent will change very rapidly throughout a period of several minutes ; if the minimum apparent separation of centres were $57''{\cdot}3$, a displacement of the Moon of $1''$, perpendicular to the line of centres, would change the position angle of this line (and therefore of the whole crescent) by 1°, and, since the mean motion of the Moon is over half a second of arc per second of time, the P.A. would change by half a degree per second, at its fastest, if the Moon's motion were fully operative. The apparent relative motion of the Moon past the Sun is somewhat less than the actual lunar motion, because the observer himself is being carried eastwards by the Earth's rotation; moreover, a separation of centres as small as $1'$ may in some cases produce a crescent too short, or too long, for satisfactory direct measurement of its P.A., or even prevent the appearance of any crescent at all; however, even allowing for these considerations one can in most cases count on a swing at the rate of 1° in 2 to 6 seconds at mid-eclipse.

The question of using this rapid run to obtain a precise correction to the Moon's place (relative to that of the Sun) attracted the writer's attention late in 1947; an important advantage is that if an accurately timed run in P.A. can be obtained, it will give corrections not merely to one coordinate but to both. It is qualitatively

clear that the smaller the minimum separation of centres the more rapid the swing, and vice versa, so that not only does the coordinate parallel to the motion come out of the mean difference between predicted and observed position angles, but also the coordinate at right angles to the motion comes out of the rate of change; both are thus obtained from P.A. measurements only; there is no need to determine any auxiliary quantity, such as the apparent separation of centres or the distance between the cusps, by measurements of absolute lengths or positions. Indeed, it has been found in practice that when corrections for the lunar limb detail are used, a value for the difference between the apparent semi-diameters of Sun and Moon is also obtained, still without any need for measurements of lengths.

The accuracy to be expected is roughly the same in both coordinates. If we take rectangular coordinates, with the axis of x parallel to the apparent relative motion of the Moon past the Sun at mid-eclipse, then, if the Sun's centre is at the origin (S, Fig. 1), the Moon's centre, L, will move along a track which, near mid-eclipse, is given by $y = \text{const.} = \kappa_0$ say.

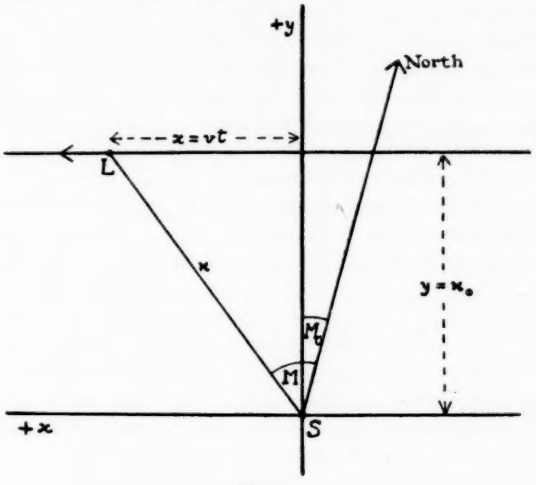

FIG. 1.

If the apparent relative velocity is v, and if the P.A. of the line SL is M, and that of the y-axis is M_0, then assuming v constant we have

$$\tan (M - M_0) = vt/\kappa_0, \tag{1}$$

where t is measured from the time of mid-eclipse. With roughly the values already assumed, i.e. $v = 0''{\cdot}32/\text{sec}$ say, $= 0'{\cdot}32/\text{min}$, and $\kappa_0 = 57''{\cdot}3$, we have $\tan (M - M_0) \cong \frac{t}{3}$ if t is in minutes; also,

$$dM = \frac{-3438}{1 + \left(\frac{vt}{\kappa_0}\right)^2} \cdot \frac{vt}{\kappa_0} \cdot \frac{d\kappa_0}{\kappa_0}. \tag{2}$$

Thus if we put $d\kappa_0/\kappa_0 = \frac{1}{573}$ and keep $v/\kappa_0 = \frac{1}{3}$, we see that the difference between the initial and final values of M changes $4'{\cdot}8$ for each $0''{\cdot}1$ change of κ_0 if the extreme values of t are $\pm 1{\cdot}5$ min, and $5'{\cdot}5$ for each $0''{\cdot}1$ if they are ± 2 min; these are increased (though not quite in proportion) if v is increased, and v may in fact often

be somewhat larger than 0′·32/min, especially towards the ends of the track. Taking the figures as they stand, they represent the effect, on the extreme values of M, of a correction to the Moon's place, in the direction perpendicular to its apparent relative motion, of 0″·1, and they are well comparable with the figure (0°·1 = 6′) by which all position angles near mid-eclipse are changed if the Moon's coordinate parallel to its motion is changed by 0″·1.

The advantage of getting both coordinates with approximately equal accuracy, from observations at a single station, is an important one; moreover, if something of the order of $2 \times 30°$ of the Moon's limb is traversed by each of the two cusps, the Moon so located will in fact be very well defined, at least provided that the limb detail at the libration in question is well known. It thus appeared very desirable to develop a method of accurately recording the rapidly changing position angle as a function of the time; timed cinematography was naturally the method demanding first study.

Since there would clearly be no difficulty in timing each of several thousand cinema-pictures to about $\frac{1}{100}$ sec, the question of the attainable accuracy reduced, in the first instance, to the question how accurately position angles could be determined from the photographic images; these must evidently be small, if they are to be obtained on 35-mm film at all, and it might be thought that the accuracy of measurement would prove to be insufficient. A little study sufficed to show that such fears were probably unfounded. It is true that one might need elaborate precautions if one attempted to determine the actual positions of the centres of Sun and Moon by measurement of the available arcs of their limbs on a small-scale picture, and that a P.A. inferred from those positions might be of low accuracy; there is, however, an alternative possibility. The line of cusps is exactly at right angles to the line of centres for two circular disks, whatever their radii, and to a first approximation there is thus no need to locate the centres themselves at all, or to know or determine the radii precisely; and if in fact one confines oneself entirely to a direct measurement of the P.A. of the line joining the two cusps, the accuracy to be expected is fairly high. With this method, incidentally, a number of disturbing factors such as (especially) irradiation and atmospheric unsteadiness tend to cancel out, by symmetry, on the average.

It is well known that the centres of photographic star-images can be located with an internal scatter of the order of half a micron or so, largely because the human eye is very sensitive to symmetry; a single cusp of a crescent is obviously a much less favourable object than a star-image, but the differential location of a pair of cusps brings in questions of symmetry again. If the photographic image be rotated, in the measuring-machine, and moved parallel to itself, until the tips of the two thin cusps just project across, or just stop short of, a single fixed fiducial thread in the field of the eyepiece, and if the cusps are really well defined, most experienced measurers would probably agree that a relative change of position, in the direction perpendicular to the thread, of $30\,\mu$ should be clearly detectable even at the fairly low magnification which must be used if the whole crescent is to be visible at one viewing. If the distance between the cusps is 10 mm, a relative displacement of $30\,\mu$ represents a rotation of about 10′ in P.A., and it is thus reasonable to expect the random scatter to remain below this figure provided the two cusps are cleanly defined. It is true that differences in their shape may introduce errors, and that such differences are in fact sure to arise on account of the rough contour of the Moon's limb; but on the average these differences also may be expected to favour

one cusp as much as the other and so to produce errors of measurement which will still be quasi-random. (The actual displacement of the cusp, which is really produced by a lunar feature, may be larger than this error in its determination, and must be allowed for directly, so far as the limb-contour is known.) A truly random scatter, even if somewhat larger than 10′, would be of little importance if two or three thousand pictures were obtained, and it thus appeared probable that the small size of the images would be no real drawback, and that the method would be worth trying out at an actual eclipse; it was decided to plan for the eclipse of 1948 November 1, visible in East Africa.

This plan was carried through, and a successful film covering nearly 3 minutes of time, with a total run in P.A. of about 76°, was obtained. The results will be discussed in detail in later papers, but it may be stated here that the above preliminary estimate of the attainable accuracy in P.A. has proved to be very conservative; in favourable cases, which include the majority of the 3 100 images secured, the actual probable error, for re-measurement of one frame by one observer, has turned out to be approximately 1′. There are some systematic effects in addition, but these are in general only of about the same magnitude, and it may even be doubted whether increasing the scale (if that were in fact practicable) would reduce them very much. Such accuracy, corresponding as it does to a differential displacement of 3μ between the two cusps, is naturally only possible if the images are sharply defined, and it was an essential feature of the work that this aim was in fact attained.

The use of partial eclipses for the purpose of determining the Moon's place is not in itself new. On several occasions it has happened that the track of a total eclipse passed fairly near to an established observatory, and programmes of positional work were then arranged for the existing equipment, especially heliometers and large refractors; in one of the earliest cases, in fact, the observatory in question (Königsberg) actually lay in the track of totality, and heliometer measurements of the P.A. and separation of the cusps gave corrections to both coordinates of the Moon's place and also to its radius, with internal probable errors of two or three tenths of a second (**1**). Measurements with ordinary equatorial telescopes (**2**) have also been made using micrometer eyepieces, and even on the projected image; all methods employed relatively few discrete determinations (photographs, drawings, or direct measurements, as the case might be). A number of other quantities besides the position angle of the line of cusps have been studied, but most of these alternatives suffer from some disadvantages; in particular, irradiation can falsify the distance apart of the cusps, the P.A. of a single cusp with respect to the centre, the "thickness" of the crescent, and (probably) the radii of curvature of the concave and convex arcs which define it (**3**). In so far as position angles are used, either alone or in combination, the effect on them of the adjustment and flexure of an equatorial mounting enters directly into the results, so that the direction of the diurnal motion must be observationally determined at the same time; sufficient accuracy in this may not be easy to get with a small number of images, especially if the atmosphere is unsteady. Some of the results obtained have, none the less, shown remarkably good internal consistency; but the work appears in general to have been regarded as a "faute de mieux" enterprise, and the method was not developed for actual expeditionary use, except for some instances (**4**) involving the partial phases as observed from inside the track of totality or annularity; the really rapid part of the run in P.A. is of course lost in that case.

Cinema equipment has also been used at eclipses, one of the earliest cases being the "limiting" eclipse of 1912 April 17, when a battery of cine cameras was set up across the track in the hope of determining just where the centre of the track lay (**5**). The records from several cameras showed Baily's beads more or less well distributed round the Moon at mid-eclipse, and the "centre" of the track was taken to lie near that camera for which the distribution was the most uniform. This method in fact (in addition to being only very rarely practicable) sacrifices much of the advantage which a cinema record is in principle well able to give, since only a small number of pictures from any one station can actually be used at all, and, in addition, only a very few points on the lunar limb, namely the bottoms of the deepest valleys, had any weight in the answer. Even for these points, corrections based on the actual lunar limb detail were not applied; the Moon was in effect defined by a circle through the centres of gravity of the beads in question, or some approximation to this, and it might be difficult to relate this "Moon" to any other one.

Much the most important application of cinematography to eclipse measurements was made by K. Kordylewski, in a programme due to T. Banachiewicz (**6**), at the eclipse of 1927 June 29. Cine cameras were set up inside the path of totality and were exposed for fairly short runs near second and third contacts. It was deliberately intended to make use of the Baily's beads which naturally appear in such a case (particularly as the eclipse was only just total), and the times of appearance and disappearance of beads, and of gaps in short residual stretches of crescent, were computed from Hayn's lunar charts and compared with observation. The internal probable error of the Moon's relative place so obtained was exceptionally small. The work was not known to the writer until after the Mombasa expedition had returned; the method evidently differs from the present one in principle, and the Moon was also still located mainly by isolated features of its limb profile and not by a continuous survey. Given an accurate knowledge of the limb at the libration in question, the fact that this knowledge is unevenly used need not of course weaken the result much; moreover, although Kordylewski was himself of the opinion that the accuracy would be less if the solar and lunar semi-diameters differed more than they did in this somewhat exceptional eclipse, it is quite possible that the loss of accuracy would in fact be slight. The method appears to deserve further trial (**7, 8**).

There does not seem to be any case of cinema work deliberately undertaken outside the track of totality; but it is here that the advantages of cinematography should probably be greatest. It is only outside the track that observations can be made at the time when the position angle is changing most rapidly; and the cusps are considerably blunter than in Kordylewski's method. Also, a small image is more acceptable for measurements of position angle than for the detailed study of beads, though the degree of detail obtained by Kordylewski was undeniably good. The very large number of pictures obtained in the present method, coupled with the unexpectedly high precision with which they can be measured, suggests that very detailed information about the Moon's limb can be used, if it is available.

In all photographic methods, there is evidently a great advantage in having an entirely stationary camera, with no coelostat. Quite apart from the increased simplicity and the greater rigidity which result, it is valuable to have the direction of the diurnal motion directly recorded, whenever position angles are required. In the present case (and probably in most others) there is then no need to rely on any angles measured in the field; all the necessary data are contained in the

film itself, and can be evaluated in the laboratory at leisure and checked in case of doubt. However, if a stationary camera is to be used, the focal length and the dimensions of the frame must be suitably related to the intended duration of the run. With a 35-mm camera, the frame measures about 23 mm × 17·5 mm; if the diurnal motion of the image is arranged to run parallel to the long side, and if the solar image is 9 mm in diameter, then, since the diurnal motion is a little slower (even on the equator) than one solar radius per minute, it is possible to keep the camera stationary and still have the entire solar image "on" for a period of rather more than three minutes. This will fill a standard 200-ft. roll of film if the speed is about 17½ frames/sec. The duration can in some cases be extended; if only the tips of the crescent are actually required to be "on", and if the line of cusps is nearly parallel to the short frame-edge at the start or finish of the run, even the centre of the Sun may sometimes be outside the frame, and a total run of about 4 minutes is quite possible. Moreover, if either the north limb of the Sun or the south one is unobscured throughout most of this time, one can expect to get a very reliable direct determination of the angle (ideally zero) between the long frame-edge and the direction of diurnal motion, since this direction appears as the common tangent to all the limb images obtained. (There is a slight curvature, if the declination is not zero, but this can easily be allowed for.) If position angles are measured, on individual frames, with respect to the direction of this same frame-edge, they can then at once be converted to true directions on the celestial sphere.

Although the azimuth of the camera thus does not appear directly in the results, it is none the less desirable to set it up in such a position that the direction of diurnal motion will in fact be closely parallel to the frame-edge (or to any other fiducial line which may be used instead). This is partly to facilitate adjusting the image to give a "central" run, but principally because the procedure for determining the direction of diurnal motion, from the photographs, will be to measure the distance of the limb from the fiducial line, on as many frames as seem necessary, and to divide the change in this distance by the distance moved parallel to the fiducial line; this latter distance may however be somewhat difficult to determine accurately, since the E.–W. coordinate of the north (or south) limb is of course poorly defined. The total distance of E.–W. travel is only about 14 mm on the assumptions used above, and, if it cannot be obtained to better than (say) 1 per cent, the angle cannot either; it would in that case be desirable for the angle to be so small that an error of 1 per cent in its value was unimportant. It is true that the E.–W. travel can be determined by other measurements (e.g. of the east or west limb, if present, or of the focal length), so that if the camera were in fact wrongly set up the results could still be used; there are also alternative measuring-procedures, at least if point-like fiducial marks are available; but these procedures all seem to involve more measurement. It is therefore preferable to get the direction of travel correct within at any rate half a degree, and there is no real difficulty in this.

This does not, of course, imply that the camera must be set up in a tilted position; the obvious procedure is to have the camera itself horizontal, and to mount a plane mirror in front of the lens; the image of the crescent may then be brought into place just before the start of the run by adjusting this mirror empirically, and it will in general be possible to select such an azimuth that at the time of the eclipse the image will travel horizontally. (This use of a mirror is clearly also the most convenient arrangement from the point of view of focus-tests, rehearsals, etc., since it is easier to adjust a mirror, so as to bring the image on to the front of the

camera, than to swing the whole assembly about for this purpose; the diurnal motion will not normally be horizontal in such rehearsals, but this is immaterial.) The required azimuth may be determined as follows.

In Fig. 2, S is the Sun, H–H the horizon, Z the zenith, N the point where the mirror-normal meets the celestial sphere, C the point (in the horizon) where the optical axis of the camera (in the sense lens-film) does so; z is the Sun's zenith distance and A' is the difference in azimuth, which it is required to obtain, between the Sun and the camera-axis. The figure is drawn for the case that the observer is looking "out" at the celestial sphere, that the hour angle and latitude are positive, and that the Sun would be south of the zenith at noon, and S'S is the direction of diurnal motion, differing from true westwards by the angle Y (always $<5'$) which arises from the Sun's changing declination. (Evidently Y is given by $\tan Y = \dfrac{\sec\delta \,.\, \Delta\delta}{15\,\Delta t}$, where $\Delta\delta$ is the change in (geocentric) declination in the time Δt; if $\Delta\delta$ is in seconds of arc, Δt is in seconds of time.) C'C is the direction of motion of the Sun's image, and it is required that this shall be horizontal, as shown. Q is the parallactic angle; it appears simpler to use Q, z and A' than to work in terms of hour angle, declination, latitude, and the absolute azimuth of C.

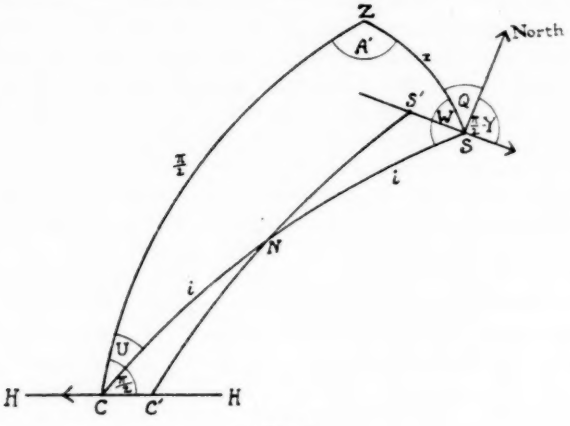

FIG. 2.

From the laws of reflection, it is clear that SNC and S'NC' are great circles, and that the two spherical triangles NCC' and NSS' are equal in all respects; NS = NC = i is the angle of incidence (and reflection) at the mirror. If W is the auxiliary angle ZSC, then from consideration of the angles at S we have

$$\mathrm{NSS'} + \pi = W + Q + \frac{\pi}{2} - Y,$$

while the angles at C give

$$\mathrm{NCC'} = \frac{\pi}{2} - U$$

if U is the angle ZCS.

The congruence of the two triangles thus gives

$$\pi - U = W + Q - Y$$

or, writing

$$Q - Y = V, \tag{2}$$

$$\sin U = \sin (W + V).$$

From the triangle ZCS we have directly

$$\sin U = \sin W \sin z,$$

so that

$$\tan W = \frac{\sin V}{\sin z - \cos V}.$$

From the same triangle we also have directly

$$\tan W = -\tan A' \sec z,$$

and eliminating W we thus obtain

$$\tan A' = \frac{\sin V \cos z}{\cos V - \sin z}. \tag{3}$$

The sign of A' in this is positive as drawn in the figure, i.e. when A' is positive, C is to the *left* of S; it will be found in fact that the meridian always lies between ZS and ZC.

If the line of fiducial marks (or the edge of the frame) is not horizontal when the base of the camera is, or if a slanting run is desired for any other reason, the formula for tan A' may still be used by writing $V = Q - Y - T$, where T is the slant of the line in question, reckoned positive if the image is to rise with advancing t.

For the special case of the 1948 eclipse in East Africa the circumstances were unfavourable; the values of cos z and cos V (or cos Q) were both fairly small, and the value of A' would thus have approached 180°. Since cos $2i = \sin z \cos A'$, $2i$ must also have approached 180°, so that an undesirably large angle of incidence on the mirror would have been involved. The altitude of the Sun at mid-eclipse was in fact only about 22° even on the coast, and the direction of diurnal motion was almost vertical, so that one could hardly expect to produce horizontal motion of the image, in a horizontal camera, by one reflection at an acceptable angle. In principle, the solution might have been to lay the camera on its side, so as to get the long edge of the frame vertical (i.e. to put $T = 90°$); but the cameras available did not appear to have been designed for use in that position and it was considered unsafe. There seemed no objection, however, to tilting the camera forward; the reflected ray would then run up at a fairly steep slant, and it was qualitatively obvious that this would ease the situation greatly. On repeating the above analysis, but with the zenith distance of the camera axis taken as w, instead of $\pi/2$, formula (3) is replaced by

$$\sin A' \cdot \frac{\cos V - \sin z \operatorname{cosec} w}{\cos z \sin V} = \cos A' - \cot w \tan z. \tag{4}$$

This equation is evidently less convenient than the other; it can, however, be solved for A' fairly readily by successive approximations, or it may be solved directly by writing it in the form $\cos (A' + a) = \cot w \tan z \cdot \cos a$, where

$$\tan a = \frac{\cos V - \sin z \operatorname{cosec} w}{\cos z \sin V};$$

in either case, a value of w is first selected more or less arbitrarily, and the two factors $(\cos V - \sin z \operatorname{cosec} w)/\cos z \sin V$ and $\cot w \tan z$ are computed for this. The formula for cos $2i$ is readily amended, and, if the final result does not give a small enough value of i, a fresh value of w may be tried; in practice, however,

it seems unlikely that a case less favourable in this respect than the 1948 one could arise, and the value then selected for w, namely 45°, is probably small enough for all circumstances. It led to practical arrangements which were in most respects quite convenient.

Choice of apparatus.—If the solar image is to be about 9 mm in diameter, the focal length must be roughly 96 cm. Cinema lenses of this focal length are not common, and would not in fact be very suitable; but since the ends of the 23-mm field will be only about 40′ from the centre, a good quality small-transit lens (i.e. a visual telescope lens corrected for coma) will give adequate cover even though it is only a simple achromat. Such a lens will normally have an aperture of 7 or 8 cm; it may be stopped down somewhat in order to reduce the intensity (or to improve the geometrical definition), but it is in fact undesirable to reduce the aperture much since the resolving power would then suffer; it was assumed above that the final 30 μ of the tip of a cusp would be clearly noticeable in the measuring machine, and it would seem advisable to keep the diffraction disk below this figure, so that the true appearance of the cusp, as caused by the actual lunar contour from moment to moment, may be reflected in the image as far as possible, and also so that the image may be as sharply bounded as possible. On the scale we have been considering, the radius of the first dark ring, in the image of a point-source, will be 30 μ if the diameter of the stop is 2 cm; it would clearly be well to keep the aperture larger than this. The f-ratio will thus be wider than $f/48$; indeed, the full aperture of $f/12$ appears preferable if the lens is good; and since a photoheliograph can work at an effective $f/250$, and with a very slow emulsion, and still give an image at $\frac{1}{500}$ sec exposure, it is plain that there is a very considerable amount of light to spare.

A reducing factor of nearly 20 can be obtained by using the mirror unsilvered; this has the additional advantage that there is then no risk of burning any filter or of burning the film if the shutter happens to stop in the "open" position. (It is of course necessary to eliminate the risk of reflection from the rear surface of the mirror; what one requires, therefore, is a "solar flat", or alternatively one face of a good spectrograph prism, of suitable size.) A further reduction in intensity will be obtained if a filter is used, and since fairly narrow-band green filters are commercially available one is thus in the fortunate position of being able to eliminate the secondary spectrum of a visual achromat very completely indeed. With a lens that is corrected for coma, as well as spherical aberration, approximately down to the Rayleigh limit, the secondary spectrum is much the most important remaining factor in photographic image-defects, especially with fully exposed images such as are desirable in the present case, and the improvement in photographic definition which a narrow green filter gives is very marked. Trial apparatus assembled on these lines, using Kodak "Background X" (panchromatic) film, a Wratten 57 filter mounted close to the camera, a lens from one of the small transits made by Troughton and Simms for the transit of Venus expeditions of 1874, and a "Vinten K" (obsolete Air Force pattern) cine camera (with its own lens removed), gave excellent test-images in the early summer of 1948; sunspots less than 10″ in size were very readily seen under the magnifying glass, and could be checked against the photoheliograph pictures of the same day.

In these tests the telescope tube was cut off some distance from the camera, so as to allow the solar image to be easily seen on the front of the camera and adjusted by moving the mirror (prism); it was found that with the green filter

in place, and using exposures of about $\frac{1}{600}$ sec, there was very little fogging by direct light from the sky or the surroundings, even when the camera was left completely uncovered after adjusting the image. This was in fact regarded as unfortunate, since it had been intended to use weak fogging to show up the edge of the frame, so that it could serve as the fiducial line in the measurements; it now appeared almost certain that when the general level of illumination was reduced by a factor of 40 or so by the eclipse, the scattered light would be decidedly too faint for the purpose. Actual fiducial marks were thus desirable, but no practicable system for impressing these on the film could be devised in the time available, and it was decided to rely on the perforations, which are punched with extraordinarily high accuracy (as indeed they must be if the projected image is not to flicker). It was realized that under these conditions a scatter could arise, both in the position angles as referred to the holes, and also in the positions themselves (which are used for the zero of position angles), if the film ran in any way variably in the camera; but it was thought that such effects would be random, and also that they must already have been shown to be small by the absence of appreciable flicker. Even if these expectations were disappointed (as, in fact, they proved to be), it would not interfere with the judgment of the method's suitability otherwise; fiducial marks could be incorporated in any later work, though no satisfactory method of actually doing this was evident in 1948.

Choice of station.—The site of the observation station will influence the shape of the crescent, the sharpness of its cusps, and the rate at which it will rotate in position angle. These quantities may be evaluated, with ample accuracy for the preliminary purpose of choosing a station, as follows.

As is usual in eclipse work, we pick a fundamental plane perpendicular at all times to the axis of the shadow; for present purposes, it is more convenient to think of the plane through the observer, satisfying this condition, than of the parallel plane through the centre of the Earth; the numerical values of m, M, n, and N are not affected by this change. In the standard notation (m, M) is then the vector giving the distance and direction from the observer to the point where the shadow-axis cuts the plane, and (n, N) is the vector giving the relative velocity of that point with respect to the observer; directions are reckoned from the projection of the Earth's axis on to this plane, in the sense north – east. Mid-eclipse is evidently (as always) defined by $M = N \pm \pi/2$, m being then a minimum equal to m_0 say, and we compute M_0, m_0, N_0, n_0 for this instant in the standard manner. Since, now, the unit of length for m and n is the Earth's equatorial radius, the apparent separation of centres of Sun and Moon, κ say, is

$$\kappa = (\pi_{\leftmoon} - \pi_{\odot})m, \tag{5}$$

where $\pi_{\odot}$ is the Sun's (geocentric) parallax, and $\pi_{\leftmoon}$ is the Moon's geocentric parallax increased to allow for the fact that the observer's distance to the Moon is appreciably less than the geocentric distance; with sufficient accuracy we may write (**9**)

$$\pi_{\leftmoon} = \pi_{\leftmoon}^{\circ}(1 + \cos z \,.\, \sin \pi_{\leftmoon}^{\circ}), \tag{6}$$

where $\pi_{\leftmoon}^{\circ}$ is the geocentric value and z the zenith distance; for present purposes we may also put $\pi_{\odot}$ and $\pi_{\leftmoon}$ constant, so that κ becomes simply proportional to m.

A little reflection makes it clear that M, which is defined in effect as a "position angle" on the fundamental plane, is also the position angle, on the celestial sphere, of the line apex–Sun–Moon as seen by the observer (looking at

the apex); or, since the convergence of meridians is entirely negligible as between apex and Sun, M is the position angle of the Moon, reckoning from the Sun. It follows, if κ is proportional only to m, that (n, N) is parallel in direction and proportional in magnitude to the vector rate of change of the vector $\boldsymbol{\varkappa}$, i.e. to the apparent relative motion of the Moon past the Sun; in fact

$$\boldsymbol{v}=(\pi_{\mathbb{C}}-\pi_{\odot})(n, N) \tag{7}$$

is this velocity, in the same angular units in which π is measured. The relative velocity (n, N) of the shadow-centre on the fundamental plane is very nearly constant; the main portion of it is the shadow's own motion, which changes only very slightly indeed in the few minutes involved, and even the secondary part (the projection on to the plane of the reversed diurnal motion of the observer) does not vary much. Accordingly, we also have $\boldsymbol{v}=\text{const.}$, as was assumed in Fig. 1; and since the axes of that figure were already chosen parallel to the directions given by N_0 and M_0 respectively, we may confine ourselves to the magnitudes of the vectors again, and write

$$\pm y=(\pi_{\mathbb{C}}-\pi_{\odot})m_0=\kappa_0=\text{const.}, \tag{8}$$

$$v=(\pi_{\mathbb{C}}-\pi_{\odot})n_0=\text{const.},$$

$$x=vt,$$

$$\tan(M-M_0)=x/y=\pm\frac{n_0}{m_0}t, \tag{9}$$

$$\kappa=\kappa_0\sec(M-M_0), \tag{10}$$

where t is in minutes of time; x, y and κ are in minutes of arc if $\pi_{\mathbb{C}}$ and $\pi_{\odot}$ are. There is no need to go through the somewhat lengthy calculations of M, m, κ, etc., for any other times; these equations give all the accuracy necessary and with very much less labour, though of course the rigorous ones must be used in the actual analysis, and will indeed naturally be employed as a check as soon as the station has been definitely selected. The double sign in (8) and (9) arises from the fact that m is always taken as positive by definition; y will be negative, in Fig. 1, and $\tan(M-M_0)$ will decrease with time, if the station is north of the centre-line; and conversely. (The opposite holds in those rare eclipses where the track crosses the Arctic or Antarctic "beyond" the pole.)

The total angle swung through, $2(M-M_0)$, is immediately determined by (9) as soon as the limiting values of t are given; t is limited by the focal length, and cannot in fact greatly exceed $\pm 2^{\mathrm{m}}$, whatever the focal length and field, without making a more elaborate lens necessary. The exact limiting times will depend also on the P.A. of the cusps at the times in question, and on the Sun's declination, but in any one eclipse they cannot vary much once the lens is chosen. Since, now, n_0 is insensitive to the observer's exact position, so long as the general neighbourhood is fixed, while m_0 is markedly sensitive, equation (9) shows that in effect the limit to the total swing that can be obtained is set by m_0; from this point of view the smaller m_0 can be made the better.

It is however clear that if κ is too small, an unsatisfactory crescent may result. If ν and σ are the apparent semi-diameters of Moon and Sun, then in an eclipse which is total on the centre-line (i.e. $\nu>\sigma$), there will be no crescent at all unless the station is chosen so that $\kappa>\nu-\sigma$, and for values only slightly above this limit the crescent will still be short, and with strongly diverging and very sharp cusps.

The question of their sharpness must first be settled; it is not entirely clear *a priori* that a separation of centres as small as 1′ might not always involve such sharp cusps that the irregularities of the Moon's outline would make them behave rather wildly, showing incipient or even complete Baily's beads for a large fraction of the total time. The two tangents at a cusp being at right angles to the radii there, it is evident that if $\kappa = 1'$, and $\nu \approx \sigma \approx 16'$, the cusp-angle will not be more than about $3\frac{1}{2}°$, and will indeed fall below even this figure if the crescent is markedly less (or more) than a semi-circle; and the slopes of mountains on the Moon being notoriously much steeper than $3\frac{1}{2}°$ in general, it might be thought that such thin cusps would be quite unmanageable. In fact, however, this is not the case; large-scale photographs of the Moon, visual observations, and the traces published by Hayn, all agree that although there are some fairly steep places on the limb, there really are relatively few of them, and the number of cases where a slope greater than three or four degrees persists for any appreciable distance is not in itself large enough to cause serious anxiety. The reason for this relative smoothness of the limb is that at the limb, if the region is a mountainous one, we are looking at a jumble of mountains one behind the other, and if one mountain falls steeply, it will often merely unmask the next in the line of sight, so that the limb is little affected. It would be increasingly unsatisfactory to sharpen the cusps still further, since even when Baily's beads did not actually form the cusps would become very sensitive even to moderate slopes on the limb; but it seemed probable that provided the minimum cusp-angle was not less than 3° one might expect serious difficulty only over short isolated stretches. For the first trial, in 1948, it was decided to accept this, and the results did in fact confirm expectation in this respect.

In Fig. 3, S and L are the centres of Sun and Moon as before, and C is a cusp; the cusp-angle is θ, and the half-lengths of the two circular arcs which limit the crescent are defined by the angles ψ and χ at the centres of Moon and Sun respectively. Then evidently

$$\cos\psi = \frac{\nu^2 + \kappa^2 - \sigma^2}{2\kappa\nu}, \qquad (11)$$

$$\cos\chi = \frac{\nu^2 - \kappa^2 - \sigma^2}{2\sigma\kappa}, \qquad (12)$$

and

$$\cos\theta = \frac{\nu^2 - \kappa^2 + \sigma^2}{2\nu\sigma}. \qquad (13)$$

From the last of these, we see at once that if κ is varied (either by moving the station, or by the course of the eclipse), $\cos\theta$ always moves in the opposite sense; since θ is always positive (and small), it follows that the smaller κ is, the smaller θ must be also. Once the station is fixed, θ will therefore have a minimum at mid-eclipse, and in estimating whether a station will permit a suitable value of θ it is sufficient to consider mid-eclipse only. For this purpose it is convenient to use $\sin\theta = (\kappa/\nu)\sin\chi$ instead of (13); it will appear below that χ (which is, of course, equal to $\psi + \theta$ or to $(\psi + \chi)/2 + \theta/2$) should probably not be allowed to fall as low as 70°, so that $\sin\chi$ will exceed 0·95 and will moreover vary only slowly as κ changes; the value of θ can in that case be estimated at once for any given κ, by writing

$$\sin\theta \cong 0{\cdot}96\,\kappa/\nu. \qquad (14)$$

ψ and χ will also have minima at mid-eclipse, but in the case of ψ it is not immediately clear that the minimum will be the smallest value since ψ has also a

maximum, namely when $\kappa^2=\nu^2-\sigma^2$, i.e. when $\chi=\pi/2$. In practice, however, it will be found that unless ν and σ are equal within a few seconds, this maximum will occur further from mid-eclipse than the run will extend, and the minimum value will usually be the smallest also; if values for some other time than mid-eclipse are nevertheless required, they can always be readily obtained by evaluating the appropriate κ from (10), and inserting it in (11).

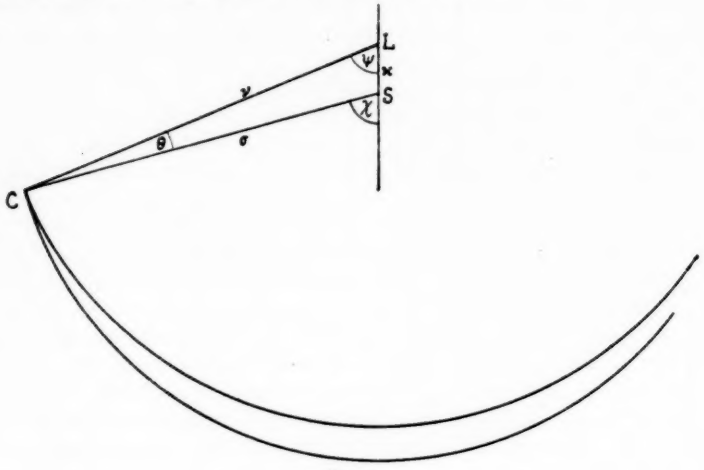

FIG. 3.

What we shall in fact be most interested in is not ψ but $(\psi+\chi)/2$, since this is the real measure of the divergence of the cusps. If these are not parallel to one another, they will not be perpendicular to the cross-hair in the measuring machine, and although we can (and, indeed, must) accept some obliquity in this respect it would probably be undesirable to allow a cusp to make an angle of less than about 70° with the cross-hair. Indeed, the risk of systematic error due to residual astigmatism in the observer's eye is considerably more serious than is sometimes realized, and, if the cusps diverge (as they would in that case) by more than 40° from each other, even the use of several different positions for every measurement may not altogether eliminate bias.

The "direction" of a cusp may be taken as perpendicular to the bisector of the angle θ, and this bisector evidently makes an angle $(\psi+\chi)/2$ with the line SL. The (total) divergence, D say, is therefore $D=\pi-(\psi+\chi)$, and we require a convenient expression for the behaviour of D as κ is varied.

Since $\sin\chi=(\nu/\sigma)\sin\psi$, we have $\cos(\psi+\chi)=\cos\psi\cos\chi-(\nu/\sigma)\sin^2\psi$, and, as $\sin^2\psi$ is readily available from (11), we obtain after a little elementary reduction

$$\cos(\psi+\chi)=-1+2\left(\frac{\nu-\sigma}{\kappa}\right)^2\left[1-\frac{\kappa^2-(\nu-\sigma)^2}{4\nu\sigma}\right]$$

or

$$\cos^2\frac{\psi+\chi}{2}=\left(\frac{\nu-\sigma}{\kappa}\right)^2\left[1-\frac{\kappa^2-(\nu-\sigma)^2}{4\nu\sigma}\right]. \tag{15}$$

Now $8\nu\sigma\approx 2\,000$, and κ^2 will seldom be as large as 20; we may thus write

$$\sin\frac{D}{2}=\cos\frac{\psi+\chi}{2}\approx\frac{\nu-\sigma}{\kappa} \tag{16}$$

with high accuracy.

Accordingly, the minimum allowable value of κ necessary to keep the (total) divergence D below any specified figure can be readily estimated in any given eclipse (i.e. for any given value of $\nu-\sigma$) from the following table:—

TABLE I

D	45°	40°	35°	30°	25°	20°	15°
$\kappa/(\nu-\sigma)$	2·61	2·92	3·26	3·86	4·62	5·76	7·66

Before evaluating κ from this table it is necessary to increase the geocentric value of ν by the same factor, $(1+\cos z \sin \pi(\circ))$, as was applied to $\pi(\circ)$. For an average eclipse, this will increase ν by about 1 per cent; in an extreme case (perigee eclipse in the zenith) the augmentation can amount to 1·8 per cent, i.e. 0′·30. It is clear that the increase in $\nu-\sigma$ may occasionally be as large as $\nu-\sigma$ itself, or even larger; there is no objection to this, but it underlines the importance of the correction.

A rough survey of the frequency with which different values of $\nu-\sigma$ occur in total eclipses indicates that in about half the cases $\nu-\sigma<0'{\cdot}4$ before applying the above augmentation to ν, or say $<0'{\cdot}6$ after applying it; even for the worst of these cases, Table I shows that if D is allowed to reach 40°, κ can be brought down to about 1′·7, which corresponds to a total change of M of over 50° if t can reach $\pm 2^{m}$ and if v is 0′·4/min. Such cases may perhaps all be classed as "favourable", in contrast to those in which larger values of κ_0 must be accepted. Geocentric values of $\nu-\sigma$ as large as 0′·85 are still fairly common; even if near-zenith cases are avoided (by picking stations towards the ends of the track, when necessary), κ will then have to be nearly 3′, and M will therefore change rather slowly; the value of such eclipses, or alternatively the dangers involved in accepting a larger value of D than 40°, would probably have to be ascertained by trial. It might be found, in practice, that κ_0 could be still further increased, since possibly the larger values of θ which would then occur would more than compensate for the small value of n_0/m_0, by diminishing the effect of lunar limb-irregularities.

If we combine the two conditions we have been considering, namely that $\sin\theta$ shall not be less than 0·060 say ($\theta \approx 3\frac{1}{2}°$) and that D shall not exceed 40°, we obtain

$$0{\cdot}060 = \sin\theta = 0{\cdot}95\kappa/\nu = 0{\cdot}95 \times 2{\cdot}92(\nu-\sigma)/\nu,$$

and putting $\nu \approx 16'$ we have roughly

$$\nu-\sigma = \frac{0{\cdot}060 \times 16'}{0{\cdot}95 \times 2{\cdot}92} = 0'{\cdot}35;$$

if $\nu-\sigma$ is less than this, the lower limit to κ is set by the sharpness of the cusps, and if $\nu-\sigma$ is greater it is set by their divergence. In the Mombasa case, $\nu-\sigma$ was a little over 0′·2, which is considerably better than an average value; the maximum value of D was 30°, and the minimum value of θ was actually $3\frac{1}{4}°$; this, coupled with the high value of v which the low altitude involved, led to a rate of rotation (31′/s at mid-eclipse), which is about the maximum obtainable, with the above restrictions. However, it is possible that a slower rate of rotation, with blunter cusps, would have been at least as satisfactory, so that eclipses in which $\nu-\sigma$ was larger would be very little inferior to the Mombasa one. (For all eclipses in which the rate of rotation is limited by θ, and not by D, it is insensitive to $\nu-\sigma$.)

The considerations in the case of an annular eclipse are in general very similar to those for a total one, and may readily be derived from them. Fig. 4 differs

from Fig. 3 only in having $\nu<\sigma$ instead of $\nu>\sigma$, and the only changes which result in the conclusions are the following:

If the eclipse is annular on the centre-line, $\sigma>\nu$, and if a crescent (and not an annulus) is to be visible, the station must be chosen so that $\kappa>\sigma-\nu$; the cusps will then always converge, instead of diverging, but otherwise what has already been said about them stands. Equations (11), (12) and (13) are not altered; $\cos\chi$ is always negative, and χ is still always greater than ψ; χ and ψ both approach π as κ approaches $\sigma-\nu$, and they both have maxima at mid-eclipse (though θ still has a minimum there); χ also has minima ($=\pi/2+\theta$), if κ is increased until $\psi=\pi/2$. ψ should probably be kept below 110°, and θ can be readily estimated from

$$\sin\theta=\frac{\kappa}{\sigma}\sin\psi\approx 0{\cdot}96\kappa/\sigma. \tag{17}$$

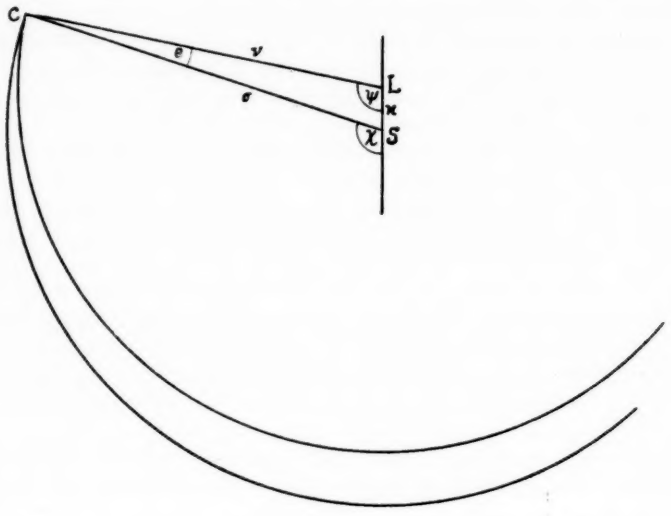

FIG. 4.

The convergence of the cusps is $-D=\psi+\chi-\pi$ and the same formula holds as before for $\sin D/2$; Table I is unaltered numerically, but the numbers now refer to $-D$ and $\kappa/(\sigma-\nu)$ respectively. The augmentation of ν now therefore allows smaller values of κ than the geocentric ν would admit; in limiting cases, as is well known, the augmentation may change the sign of $\sigma-\nu$ and so produce an eclipse which is total in the central part of its run and annular only at the ends.

The statistics of annular eclipses appear at first sight somewhat unfavourable by comparison with those of total ones, since the geocentric values of $\sigma-\nu$ not infrequently exceed 1′·2 and occasionally exceed even 1′·5; however, if we consider also the fact that annular eclipses happen oftener, the worst cases lose significance. If we again allow an average augmentation of 0′·2, we find that for the best of the total and annular cases, respectively, the probability of getting an augmented value of $\nu-\sigma$ not exceeding 0′·6 is about $\frac{1}{3}$ per year, and the probability of getting a diminished value of $\sigma-\nu$ in the same range approaches $\frac{1}{2}$ per year. Exact figures depend on the actual augmentation in each case, and this is partly a matter of the smallest zenith distance reached by the eclipse, and partly a matter

of how far from the ends of the track the station is placed. (The augmentation is naturally negligible at the ends of any track, whether short or long.) Annular eclipses do of course also involve a smaller relative velocity of the Moon, on the average, and also a smaller value of $\pi_{\mathbb{C}}$; so that on balance there will be nearly equal numbers of favourable cases in either class. The total number of total and annular eclipses is roughly double the number of "favourable" ones in the above sense, and, as we have seen, it may well be found that even the "unfavourable" ones can give good results, owing to the relatively blunt cusps which occur if a slow rate of rotation is accepted. Thus an eclipse programme devoted to this method, if it can in fact use most annular eclipses as well as all total ones (so far, that is, as questions of accessibility and weather permit), should have many more opportunities of being carried out than those programmes which are restricted to total eclipses. Whenever the track runs over, or close past, two regions which are far apart on the Earth's surface, the method will have geodetic applications, and the fact that two coordinates, and not merely one, are obtained will obviously be in its favour in this respect also. The fact that sites may be picked on either side of the centre-line will sometimes greatly widen the practical possibilities, especially when the actual track just fails to touch land over an important part of its length. The personnel requirements are modest, since in principle one man can do all that is required at any one station.

A detailed account of the actual work at Mombasa, and an analysis of the results obtained, will be given in subsequent papers.

Royal Observatory,
Greenwich:
1952 *November* 4.

References

(1) M. Wichmann, *A.N.*, **33,** 309, 1852.

(2) H. Kobold, *Strasb. Ann. II*, Annex A, 1899 (Internal mean errors mostly $<0''.2$); V. Knorre, *A.N.*, **152,** 115, 1900; E. Hartwig, *ibid.*, 391; H. Struve, *A.N.*, **153,** 299, 1900; J. Merlin, *C.R.*, **144,** 20, 1907; F. Hayn, *A.N.*, **193,** 117, 1913, and **201,** 185, 1915; H. Naumann, *A.N.*, **193,** 123, 1913, and **201,** 217, 1915; Leiden Obs. Staff, *B.A.N.*, **4,** 71, 1927; S. Ishii, *Jap. J. Astr. Geoph.*, **11,** 37, 1933, and *T.A.B.*, **25**; and many others, some of comparable accuracy.

(3) Systematic differences, more or less irreconcilable with the formal probable errors, were found between the results obtained with different methods, by C. le Morvan, *Bull. Astr.*, **32,** 337, 1915; Leiden observers, *loc. cit.*; C. Wirtz, *A.N.*, **239,** 121, 1930; also there are a number of cases in which the results differ from the accepted values by large amounts, although the "probable errors" were small.

(4) For example: E. Stephani, *A.N.*, **191,** 437, 1912; R. Schorr, *ibid.*, 429; E. Perepelkin, *A.N.*, **232,** 127, 1928; C. Wirtz, *loc. cit.*; J. E. Willis, *A.J.*, **47,** 109, 1939; and some others.

(5) Da Costa Lobo, *C.R.*, **154,** 1396, 1912; *cf.* also A. de la Baume-Pluvinel, *ibid.*, 1139; E. Carvallo, *ibid.*, 1072, 1142; F. Vlès, *C.R.*, **155,** 545, 1912; R. Schorr, *loc. cit.*; K. Graff, *A.N.*, **192,** 43, 1912, and *Sitzb. Ak. Wien* (IIa), **147,** 271, 1938.

(6) K. Kordylewski, *A.A.* (b), **1,** 133, 1932.

(7) It was used in 1936 (T. Banachiewicz, *A.A.* (c) **3,** 16, 1936); in 1945 and 1947 (see H. Kristenson, *Stockh. Obs. Ann.*, **17,** pp. 9, 10, 1951); and in a modified form in 1948 (see T. Panay, *Trans. Amer. Geoph. Union*, **31,** No. 6, 1950); detailed results do not yet seem to be available.

(8) Some other methods, not considered in the text because they differ from the present one in both principle and technique, are nevertheless important; especially Hayn's photographic method (F. Hayn, *A.N.* **233,** 183, 1928), and Lindblad's spectrophotometric one (H. Kristenson, *loc. cit.*, pp. 3–96).

(9) R. d'E. Atkinson, *M.N.*, **111,** 450, 1951.

ON THE ORIGIN OF INERTIA

D. W. Sciama

(Received 1952 August 20)*

Summary

As Einstein has pointed out, general relativity does not account satisfactorily for the inertial properties of matter, so that an adequate theory of inertia is still lacking. This paper describes a theory of gravitation which ascribes inertia to an inductive effect of distant matter. In the rest-frame of any body the gravitational field of the universe as a whole cancels the gravitational field of local matter, so that in this frame the body is "free". Thus in this theory inertial effects arise from the gravitational field of a moving universe. For simplicity, gravitational effects are calculated in flat space-time by means of Maxwell-type field equations, although a complete theory of inertia requires more complicated equations.

This theory differs from general relativity principally in the following respects:

(i) It enables the amount of matter in the universe to be estimated from a knowledge of the gravitational constant.

(ii) The principle of equivalence is a consequence of the theory, not an initial axiom.

(iii) It implies that gravitation must be attractive.

The present theory is intended only as a model. A more complete, but necessarily more complicated theory will be described in another paper.

---

1. *Introduction.*—In this paper we construct a tentative theory to account for the inertial properties of matter. These properties imply that at each point of space there exists a set of reference frames in which Newton's laws of motion hold good—the so-called "inertial frames". If other frames are used, Newton's laws will no longer hold unless one introduces "fictitious" (inertial) forces which depend on the motion of these frames relative to an inertial frame.

The question then arises: what determines the inertial frames? Newton asserted that they were determined by absolute space. However, absolute space is not observable in any other way, and it has been suggested that it is more satisfactory to attempt to correlate the inertial frames with observable features of the universe. In particular, Berkeley (**1**) and Mach (**2**) maintained that inertial frames are those which are unaccelerated relative to the "fixed stars", that is, relative to a suitably defined mean of all the matter in the universe. This statement is usually known as Mach's principle. As this principle will be used as a guide in constructing our theory, we shall first discuss its general implications.

The view that the problem of motion can be completely discussed in terms of observables implies that a kinematical description of all the relative motions in the universe completely specifies the system, so that kinematically equivalent motions must be dynamically equivalent. For instance, the statement that the Earth is rotating and the universe is at rest should lead to the same dynamical

* Received in original form 1952 January 3.

consequences as the statement that the universe is rotating and the Earth is at rest, whereas this is not true in a scheme based on absolute space. Using Mach's principle we can predict that the angular velocity of the Earth, as deduced from a local dynamical experiment (such as the motion of a Foucault pendulum), will be the same as that deduced kinematically from the apparent motion of the fixed stars. This prediction cannot be made in Newton's theory, because there is then no causal connection between the motion of the stars and the existence of inertial forces at the Earth; the two observations give the same result only because, as it happens, the stars are not rotating relative to absolute space.

If the rest of the universe determines the inertial frames, it follows that inertia is not an intrinsic property of matter, but arises as a result of the interaction of matter with the rest of the matter in the universe. This immediately raises the problem of how Newton's laws of motion can be so accurate despite their complete lack of reference to the physical properties of the universe, such as the amount of matter it contains. It was largely this problem which originally prevented the general acceptance of Mach's ideas, and one of the requirements of a theory of inertia that is consistent with Mach's principle is that it should account for the *apparent* irrelevance of the properties of the universe.

The observed fact that a gravitational force is *locally* indistinguishable from an inertial force, in that each induces the same acceleration in all bodies, suggested to Einstein that it is the gravitational influence of the whole universe which gives rise to inertia. General relativity was devised to incorporate this idea, but, as emphasized by Einstein (**3, 4,** *cf.* **5,** p. 97), it failed to do so. Einstein showed that his field equations imply that a test-particle in an otherwise empty universe has inertial properties. In view of this it seems to be worth while searching for theories of gravitation which imply that matter has inertia only in the presence of other matter. In this paper we describe what appears to be the simplest possible theory of gravitation that has this property, though this theory is incomplete in other respects.

2. *General formalism.*—Our problem is to construct a formalism in which the motion of a body is influenced by the presence of other bodies, but in which the concepts of "inertia" and "inertial frames" do not have to be introduced *a priori*. We shall represent the influence of the bodies on each other by a set of quantities defined at all points of space and time. As we are ignoring electromagnetic effects in this paper, we say that these quantities describe the gravitational field. The field is determined by the bodies (the sources) by means of a set of differential equations, together with suitably chosen boundary conditions. These equations show how the field can be determined from the motion (and other properties) of the sources.

In addition we must know how the field affects the motion of the sources. For this purpose we introduce the following postulate: *in the rest-frame of any body the total gravitational field at the body arising from all the other matter in the universe is zero.* In Newtonian language we could say that the universe moves relative to any body in such a way that the body never experiences a force—the difference from the ordinary Newtonian theory being that the forces acting on the body are here derived entirely from the matter in the universe.

We must now set up the equations relating the gravitational field to its sources. In the rest-frame of any particle we assume the field to be derivable from a potential in Minkowski space, that is, we do not describe it in terms of a curved

space. Kinematical considerations (Section 6) show that the potential should be a tensor of the second rank, but the use of such a potential leads to rather involved mathematics which tends to obscure the physical significance of the theory. It seems advisable to begin by working with the simplest mathematical scheme which contains the physically important aspects of the problem. The simplest type of potential we could use is a scalar, but as we shall see (Section 4 (ii)), this would not give rise to inertia. The next simplest possibility is a vector potential, and with a theory based on such a potential we can reproduce the main properties of inertia. In this paper we shall confine our attention to such a potential. This simplification is useful for gaining insight into the problem, but naturally it has its limitations, some of which are mentioned in Section 6. The more elaborate equations that are needed for a tensor potential will be described in a subsequent paper (hereafter called II).

In a theory based on a vector potential, the field is an antisymmetrical tensor —the curl of the potential. The only linear second-order differential tensor equations for a field of this type that imply the conservation of source are (**6**) Maxwell's equations, which accordingly we shall adopt. We emphasize that although our equations have the same formal structure as Maxwell's, they describe purely gravitational effects, electromagnetic phenomena being outside the scope of this paper.

In order to apply the theory to even as simple a problem as the motion of a particle in the gravitational field of the Earth, we must know the distribution of matter in the universe. In practice we shall have to approximate to this distribution in some way. The type of approximation that will be most useful depends on the relative importance of near and distant matter. Since the amount of matter at a given distance increases roughly with the square of the distance, it follows that if the influence of matter falls off more slowly than the inverse square of the distance, then very distant matter is of predominant importance. It is convenient to anticipate that this is indeed the case (Section 4 (iii)). This means that a smoothed-out model of the universe should be a good first approximation, local irregularities having only a small effect which can easily be estimated. It also means that local phenomena are strongly coupled to the universe as a whole, not just to local conditions. This in turn means that local experiments, if interpreted by means of this theory, can give us information about the structure of the universe as a whole (*cf.* **5, 7**). The correctness of this information can, in principle, be tested by independent and more direct considerations.

We shall take as our smoothed-out model a homogeneous and isotropic distribution of matter of density ρ expanding (relative to any point as origin) according to the Hubble law $\mathbf{v} = \mathbf{r}/\tau$, where $\mathbf{v}$ is the velocity of matter at distance $\mathbf{r}$, and τ is a constant. This neglects certain relativistic difficulties such as the significance of velocities exceeding that of light, but for the tentative theory developed in this paper we shall not concern ourselves with these problems; a consistent treatment will be given in II. This model is one of those in which there is a natural state of rest at each point, namely, that in which the observed distribution of the red-shifts of distant matter is isotropic. Thus we can speak of a body being at rest relative to the universe.

3. *Inertia-induction.*—In order to show how inertia arises in this formalism, we consider the behaviour of a test-particle in the presence of a single body

superposed on the smoothed-out universe. The problem is to determine what motion of the system universe-plus-body relative to the test-particle makes the total gravitational field at the particle zero.

It is convenient to begin by calculating the potential at a test-particle that is at rest in a universe containing no irregularities. Since our field equations have the same form as Maxwell's, we can use electrodynamic formulae to calculate the potential, and to bring out the analogy with electrodynamics we use a similar notation and terminology, but we emphasize that in this paper we shall be concerned with purely gravitational phenomena.

Retardation effects are taken to arise in the same way as in electrodynamics, so that the contribution of any region of the universe to the potential at a point P at time t is computed by ascribing to that region just the properties that are observed at P at time t.

We thus have for the scalar potential (**8**)

$$\Phi = -\int_V \frac{\rho}{r} dV. \tag{1}$$

We use the minus sign in (1) because inertial mass then turns out to be positive, but in fact either sign can be used (Section 4(vii)). The vector potential $\mathbf{A}$ vanishes by symmetry.

We shall assume that matter receding with velocity greater than that of light makes no contribution to the potential, so that the integral in (1) is taken over the spherical volume of radius $c\tau$. An assumption of this sort is necessary since we have naïvely extrapolated the Hubble law without considering relativistic effects, and should give the correct order of magnitude. A relativistic treatment is given in II.

Since the density is supposed uniform, (1) gives

$$\Phi = -2\pi\rho c^2\tau^2. \tag{2}$$

Owing to our assumptions, the numerical factor 2π is only approximate.

We now calculate the potentials for the simple case when the particle moves relative to the smoothed-out universe with the small rectilinear velocity $-\mathbf{v}(t)$. In the rest-frame of the particle the universe moves rectilinearly with velocity $\mathbf{v}(t)$. Now at time t there will be observable at the particle, in addition to the Hubble effect, a Doppler shift corresponding to $\mathbf{v}(t)$ from all parts of the universe. Hence, in computing the potential in the rest-frame of the particle at time t, we must ascribe to every region of the universe the velocity that is observed at time t, that is, $\mathbf{v}(t) + \mathbf{r}/\tau$.

Neglecting terms of order v^2/c^2, we have

$$\Phi = -2\pi\rho c^2\tau^2$$

as before. The vector potential no longer vanishes, but has the value

$$\mathbf{A} = -\int_V \frac{\mathbf{v}\rho}{cr} dV. \tag{3}$$

Since $\mathbf{v}$ is independent of $\mathbf{r}$, we can take it outside the integral. We then obtain

$$\mathbf{A} = \frac{\Phi}{c}\mathbf{v}(t).$$

Since the change of ρ with time is very small, the gravelectric part of the field is approximately

$$\mathbf{E} = -\operatorname{grad}\Phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$$
$$= -\frac{\Phi}{c^2}\frac{\partial \mathbf{v}}{\partial t},$$

while the gravomagnetic field is

$$\mathbf{H} = \operatorname{curl}\mathbf{A} = 0.$$

So far we have been concerned with a universe that has no irregularities. We now suppose that a body of gravitational mass M is superposed on this universe and is at rest relative to it. The field of this body in the rest-frame of the test-particle is then

$$-\frac{M}{r^2}\hat{\mathbf{r}} - \frac{\phi}{c^2}\frac{\partial \mathbf{v}}{\partial t}, \tag{4}$$

where r is the distance of the body from the test-particle, $\phi(=-M/r)$ is the potential of the body at the test-particle, and

$$\hat{\mathbf{r}} \cdot \frac{d\mathbf{v}}{dt} = \frac{dv}{dt}.$$

The total field at the particle is zero if

$$-\frac{M}{r^2} - \frac{\phi}{c^2}\frac{dv}{dt} = \frac{\Phi}{c^2}\frac{dv}{dt}$$

or

$$\frac{M}{r^2} = -\left(\frac{\Phi+\phi}{c^2}\right)\frac{dv}{dt}. \tag{5}$$

This equation asserts that the system universe-plus-body accelerates relative to the test-particle at a rate determined by the mass and distance of the body. In accordance with the discussion of Section 1, we may re-interpret this result by saying that the particle accelerates towards the body, which is at rest relative to the universe, and thereby make contact with the Newtonian view-point. Indeed we obtain for rectilinear motion *a combination of Newton's laws of motion and of gravitation, with the inertial frame determined by Mach's principle.*

Furthermore, the gravitational constant satisfies the equation

$$\frac{\Phi+\phi}{c^2} = -\frac{1}{G}$$

or, since $\phi \ll \Phi$ (*cf.* Section 4 (iii)),

$$G\Phi = -c^2. \tag{6}$$

4. *Consequences of the theory.*—(i) Equation (6) implies that the total energy (inertial plus gravitational) of a particle at rest in the universe is zero. It can in principle be tested observationally, for, in conjunction with (2), it implies that

$$2\pi G\rho\tau^2 = 1$$

or

$$G\rho\tau^2 \sim 1, \tag{7}$$

since the numerical factor is only approximate.

Equation (6) implies that the gravitational constant at any point is determined by the total gravitational potential at that point, and so by the distribution of

matter in the universe. This illustrates the fact mentioned above that if local phenomena are strongly coupled to the universe as a whole, then local observations can give us information about the universe as a whole. With our assumption about the structure of the universe, a laboratory determination of G, combined with an astronomical determination of τ, enables us to deduce the mean density of matter in space (apart from the uncertainty in the numerical constant).

Taking $G = 6 \times 10^{-8}$ c.g.s. units and $\tau = 6 \times 10^{16}$ sec we have from (7)

$$\rho \sim 5 \times 10^{-27} \text{ g cm}^{-3}. \tag{8}$$

A mean density of this amount is much larger than the usual observational estimates ($\sim 10^{-30}$ g cm^{-3}), but these values refer only to matter condensed into nebulae, and the mean density of internebular material may well exceed the overall density of material condensed into nebulae (**5**, p. 45). Indeed, if the nebulae are supposed to have condensed from an internebular medium, there is no reason why the medium should have been exhausted in the process. It is probable that ρ has an upper limit of $\sim 10^{-25}$ g cm^{-3}, as this appears to be the mean density within some nebulae. Hence (8) is not inconsistent with observation.

(ii) We see from the argument leading to (5) that "inertia-induction" arises from the term $\partial \mathbf{A}/\partial t$, that is, from the "radiation-field" of the universe. Had we used a scalar potential such a term would not have arisen, there would have been no inertia-induction, and we would not have been able to obtain Newton's law of motion (5).

(iii) The contribution of matter to local inertia falls off only inversely as the distance, since $\partial \mathbf{A}/\partial t$ is proportional to the scalar potential (*cf.* the radiation field of an accelerating charge (**8**, p. 22)). This means that the main contribution comes from distant matter—(1) shows that 99 per cent of local inertia arises from matter further away than 10^8 light-years. The fractional contributions of the Earth, Sun and Milky Way are 10^{-9}, 10^{-8}, 10^{-7} respectively (taking the mass of the Milky Way as 10^{44} g); this justifies the neglect of ϕ in (6). According to our theory, then, local phenomena are strongly coupled to the universe as a whole, but owing to the small effect of local irregularities this coupling is practically constant over the distances and times available to observation. Because of this constancy, local phenomena appear to be isolated from the rest of the universe.

(iv) As has already been pointed out, a theory of inertia that is consistent with Mach's principle must account for the fact that Newton's methods were so successful despite their complete lack of reference to the properties of the universe. The work of Section 3 shows that our theory satisfies this requirement. The universe affects local phenomena at just the two points where Newton's work contains arbitrary elements, namely (*a*) in the choice of inertial frames, and (*b*) in the value of the gravitational constant.

(v) Relation (7) is a consequence of many general relativistic models (**5**, *cf.* also **9**) as well as of our theory. However, our theory is disproved if (6) disagrees with observation, whereas general relativity is not, since it leads to many models in which (6) is false. Hence in general relativity the observed value of the gravitational constant gives no information about the amount of matter in the universe. For instance, if the Milky Way was thought to be the sole occupant of the universe, general relativity would still be consistent with the observed value of G, whereas in our theory there would be a discrepancy of a factor $\sim 10^7$. Indeed, if we accepted Mach's principle for the reasons given

in Section 1, we would, on the basis of our theory, *predict* that the universe contained vastly more matter than had yet been observed.

(vi) The principle of equivalence asserts that the phenomena occurring in a gravitational field are the same as those occurring in the absence of the field if observations are made from a suitably chosen accelerating frame. For instance, consider an observer confined to a closed laboratory. According to the principle of equivalence the observer cannot distinguish between the following two events:

(*a*) a gravitating mass is suddenly placed near the laboratory;

(*b*) the laboratory is suddenly accelerated, e.g. by being pulled by a rope.

In our theory this is explained by the fact that in case (*b*) *the motion of the universe relative to the observer produces the same gravitational field at the laboratory as that of the mass in case* (*a*) (*cf.* the argument leading to (5)).

General relativity has difficulty in explaining the principle of equivalence because it predicts that one gravitating mass in an otherwise empty universe produces effects similar to those calculated with a full universe in Section 3. Since there is no universe to act as the source of the field in case (*b*), it is difficult to see why the principle of equivalence should be true.

(vii) It follows from our equations that gravitation is attractive, whereas in general relativity the sign of the field is not determined. This result arises because our equations imply that the field at a point P due to a mass is *decreased* in absolute magnitude if the mass has a component of acceleration towards P, and *increased* if it has a component away from it, whichever the sign of the field (*cf.* (4)). Thus the field of the universe can only cancel the field of a local mass if the universe moves so that the mass, which is at rest relative to it, accelerates towards the particle. This is why the sign in equation (1) is arbitrary: had we chosen a plus sign, gravitation would have been repulsive and inertial mass negative (corresponding to changing the sign of both sides of (5)), so that the particle would still accelerate towards the body.

5. *Uniform rotation.*—In Section 3 we assumed that the universe-plus-body moved in a straight line relative to the test-particle. We shall now derive another possible motion of the system in which the universe and body rotate with constant angular velocity about an axis through the centre of the body perpendicular to the line joining it to the particle.

In the rest-frame of the particle we set up a Cartesian system of axes with origin at the body and z-axis along the axis of rotation. If the universe were non-rotating, its gravitational potential near the origin would be

$$A_x = A_y = A_z = 0; \quad \Phi = -1, \tag{9}$$

choosing units so that $G = c = 1$. If it rotates, its potential in the x, y plane near the origin is

$$\begin{aligned} A_x &= \omega y, \\ A_y &= -\omega x, \\ A_z &= 0; \\ \Phi &= -[1 + \omega^2(x^2 + y^2)]^{1/2} \\ &= -(1 + \omega^2 r^2)^{1/2}. \end{aligned}$$

This follows from the transformation properties of four-vectors (**10**). It is only true for distances from the origin for which the potential of a non-rotating

universe differs negligibly from (9), since Rosen's result was derived for a four-velocity which has the values (9) everywhere in the non-rotating frame. This restriction is of no importance for phenomena on a galactic scale since the potential of the universe at the centre of the galaxy differs from that at the edge by less than one part in 10^{10}.

We thus get

$$\mathbf{E} = -\operatorname{grad}\Phi - \frac{\partial \mathbf{A}}{\partial t}$$

$$= \frac{\omega^2 \mathbf{r}}{(1+\omega^2 r^2)^{1/2}}$$

$$\sim \omega^2 \mathbf{r}, \quad \text{for} \quad \omega r \ll 1.$$

The field of the body, neglecting its rotation, is

$$-\frac{M}{r^2}\hat{\mathbf{r}}.$$

The total field is zero if

$$\frac{M}{r^2} = \omega^2 r. \tag{10}$$

In the rest-frame of the universe, this is the usual Newtonian equation for circular motion. In the rest-frame of the particle, however, Newton's law of motion only holds if we introduce a fictitious centrifugal force-field. Equation (10) shows that in our theory this field is not fictitious, but arises from the gravitational effect of a rotating universe, in agreement with Mach's principle.

In contrast to the case of Section 3, the gravomagnetic field is not zero (*cf.* the magnetic field of a rotating charge-distribution). In fact, we have

$$\mathbf{H} = \operatorname{curl}\mathbf{A} = 2\boldsymbol{\omega}.$$

Since the test-particle is at rest, this field has no effect; but for another test-particle constrained to move relative to the first with uniform velocity $\mathbf{v}$, there would be in its rest-frame a gravelectric field

$$\mathbf{v} \wedge \mathbf{H} = 2\mathbf{v} \wedge \boldsymbol{\omega} \tag{11}$$

acting on it. This can be seen directly by taking this particle to be at rest and determining $\mathbf{E}$ in the new frame. (*Cf.* the relativistic deduction of the Lorentz force $e(\mathbf{E}+\mathbf{v}\wedge\mathbf{H})$ from the expression $e\mathbf{E}$.)

The field (11) just corresponds to the Coriolis field of Newtonian theory, and like the centrifugal field it is here ascribed to the rotating universe. Thus, in our theory, we can regard the Earth as stationary and a Foucault pendulum as pulled around by the gravomagnetic field of the rotating universe.

Another feature of this interpretation of the Coriolis field is that the bending of light in a gravitational field is observable in the laboratory. In fact, the experiments of Sagnac (**11**) and Michelson and Gale (**12**) showed that in a rotating frame of reference light does not travel in straight lines. If we interpret this phenomenon according to our theory, we may suppose the frame to be at rest and the universe to be rotating around it. The resulting gravitational field then bends the path of the light. According to this point of view the Sagnac effect is of the same nature as the refraction of light passing through the gravitational field of the Sun. A quantitative treatment of the motion of light in a gravitational field requires, however, the more elaborate equations of II.

The results of this section enable us to reconsider the problem of absolute rotation from the standpoint of our tentative theory. In fact, rotation does not give rise to any problems not already raised by translation, but the conflict between absolute and relative motion has usually been discussed in terms of rotation. The main argument for the view that there is an absolute standard of rotation has been the fact that local experiments can detect rotation—experiments in which the rest of the universe has not appeared to play an essential part. The present theory answers this argument by showing how the rest of the universe *can* play an essential part in local phenomena. The results of local experiments can be interpreted as giving us information about the universe as a whole—an interpretation which is verifiable by direct observation.

6. *Limitations of the theory.*—We have so far derived Newton's law of motion for two special cases: rectilinear motion and uniform rotation. A satisfactory theory of inertia must, of course, derive Newton's law for all possible motions. Now it follows from the relativistic form of Newton's law (**13**) that for general motions inertial forces are derived from a tensor potential and not from a vector potential. This means that the gravitational field of the moving universe, which in our theory gives rise to inertial forces, must also be derived from a tensor potential, so that our present theory is incomplete.

Furthermore, in a theory based on a vector potential it is difficult to give a consistent relativistic discussion of the structure of the universe as a whole. It is also difficult to describe the motion of light in a gravitational field. In paper II we shall describe a theory, using a tensor potential, in which these difficulties do not arise.

I am grateful to my colleagues for commenting on the manuscript, to the referees for their valuable and detailed criticisms of the first draft of this paper, and to Mr H. Bondi and Mr T. Gold for their constant advice and encouragement.

Trinity College,
Cambridge :
1952 *August* 19.

References

(1) G. Berkeley, *The Principles of Human Understanding*, 1710; *De Motu*, 1726.
(2) E. Mach, *History and Root of the Principle of the Conservation of Energy*, 1872; *The Science of Mechanics*, 6th ed., 1904.
(3) A. Einstein, *S.B. preuss. Akad. Wiss.*, p. 142, 1917.
(4) A. Einstein, *The Meaning of Relativity*, p. 94 ff., Methuen, 1946.
(5) H. Bondi, *Cosmology*, Cambridge, 1952.
(6) H. Weyl, *Am. J. Maths.*, **66**, 591, 1944.
(7) H. Bondi and T. Gold, *M.N.*, **108**, 252, 1948.
(8) W. Heitler, *The Quantum Theory of Radiation*, Sect. 3, Oxford, 1944.
(9) G. J. Whitrow and D. G. Randall, *M.N.*, **111**, 455, 1951.
(10) N. Rosen, *Phys. Rev.*, **71**, 54, 1947.
(11) G. Sagnac, *J. de Phys.*, **4**, 177, 1914.
(12) A. A. Michelson and H. G. Gale, *Ap. J.*, **61**, 140, 1925.
(13) A. S. Eddington, *The Mathematical Theory of Relativity*, Sect. 15, 28, Cambridge, 1930.

THE MOTION OF A GAS CLOUD EXPANDING INTO A VACUUM

D. C. Pack

(Communicated by J. M. Jackson)

(Received 1952 October 2)

Summary

The motion of a semi-infinite gas cloud bordering upon a vacuum is studied by the method of characteristics. Problems previously discussed by Burgers and Copson are illustrated by diagrams; they involve gas flows which are always continuous. It is shown that unless certain conditions are satisfied by the initial state in the cloud the subsequent flow cannot remain indefinitely continuous; shock-waves must appear within the gas. An example of the breakdown of continuity is given. A singular solution is considered in which all the characteristics of one family are concurrent in a point of the (x, t) plane. Finally, it is shown that the high eventual velocity of these gas clouds, which was considered to be a physically unsatisfactory feature of the previous solutions, is not to be avoided by variation of the initial inhomogeneities.

---

1. *Introduction.*—The equations which govern the flow of an ideal gas, neglecting the effects of viscosity and heat-conduction, have recently been used for the study of the motion of interstellar clouds. These clouds consist almost exclusively of hydrogen, and the value of γ, the ratio of specific heats for this gas, may be taken as $\frac{5}{3}$. For this particular value of γ an analytical solution of the equations of motion is easily obtained.

Burgers* considered the case of a semi-infinite homogeneous gas cloud which is at rest at some instant, its front adjoining a vacuum; he gives formulae for the characteristics of the motion which instantly begins. These are the characteristics of a "simple wave" (see below). The face of the gas moves forward into the vacuum with a velocity $3a$, where a is the speed of sound in the gas cloud in its initial state. At the same time a wave moves back into the cloud with velocity a, this wave separating the part of the cloud which has begun to move from the part which remains at rest.

McVittie† took up Burgers' work and tried to extend it to less artificial initial circumstances, starting with a stationary non-homogeneous gas cloud. On the surface of the vacuum the velocity of sound c is zero, so that there is no initial discontinuity of pressure, and the value of c rises continuously towards the interior of the cloud. The discontinuity of initial velocity of the gas cloud in Burgers' work is removed by this alteration; unfortunately, however, according to the physical model which McVittie adopted, the whole of the inhomogeneous cloud is never moving at once, whereas such a cloud must begin to move immediately throughout its whole mass under any likely equation of state.

* J. M. Burgers, *K. Ned. Akad. v. Wet.*, **49**, 588–607, 1946.
† G. C. McVittie, *M.N.*, **110**, 224–237, 1950.

Copson* applied the theory of Riemann invariants to the further example of a limited region of inhomogeneity backed by a region of homogeneous gas. In the limited region the speed of sound increased continuously from zero at the vacuum surface to a, the value in the homogeneous gas, at the inner boundary.

In the following sections the solutions discussed above are illustrated by the method of characteristics, and the same method is used to discuss certain variations of the initial conditions which lead to new types of flow within the gas clouds. Conditions are given which are necessary if the flow is to remain continuous.

The method of characteristics is a general one, applicable to any value of the ratio of specific heats and adaptable to any proper formulation of the initial value problem. With knowledge as to whether a solution will or will not remain continuous, it can yield, in diagram form and without detailed mathematical analysis, the essential features of the flow; otherwise, or to obtain quantitative information, a computation is necessary. In all cases the characteristics diagram enables the nature of a solution to be seen at a glance.

It is shown that the eventual high velocity of advance of gas clouds bordering upon a vacuum, which was considered an unsatisfactory feature of the previous solutions, is not to be avoided simply by varying the initial inhomogeneities.

2. *The characteristics.*—Let u, p, ρ be respectively the velocity, pressure and density in a one-dimensional unsteady motion, parallel to the x-axis, of ideal gas expanding according to the adiabatic law $p=K\rho^{\gamma}$, K being a constant. The equations of motion and continuity are respectively

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}, \tag{1}$$

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t}+\frac{u}{\rho}\frac{\partial \rho}{\partial x}+\frac{\partial u}{\partial x}=0. \tag{2}$$

The velocity of sound in the gas is c where $c^2=dp/d\rho=\gamma p/\rho$. Writing (1) and (2) in terms of u and c, the equations become

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+\frac{2c}{\gamma-1}\frac{\partial c}{\partial x}=0, \tag{1a}$$

$$\frac{2}{\gamma-1}\frac{\partial c}{\partial t}+\frac{2u}{\gamma-1}\frac{\partial c}{\partial x}+c\frac{\partial u}{\partial x}=0. \tag{2a}$$

The equations (1a) and (2a) may be replaced, after simple manipulation, by the equivalent pair

$$\left\{\frac{\partial}{\partial t}+(u+c)\frac{\partial}{\partial x}\right\}\left(\frac{2c}{\gamma-1}+u\right)=0, \tag{3}$$

$$\left\{\frac{\partial}{\partial t}+(u-c)\frac{\partial}{\partial x}\right\}\left(\frac{2c}{\gamma-1}-u\right)=0, \tag{4}$$

which shows that

(i) along a curve in the (x, t) plane for which $\frac{dx}{dt}=u+c$, the expression $2r=\frac{2c}{\gamma-1}+u$ is constant, and (5a)

(ii) along a curve in the (x, t) plane for which $\frac{dx}{dt}=u-c$, the expression $2s=\frac{2c}{\gamma-1}-u$ is constant. (5b)

* E. T. Copson, *M.N.*, **110**, 238–246, 1950.

These curves are the *characteristics* of the differential equations governing the motion, and the conditions $r =$ constant, $s =$ constant, holding respectively along the two families of curves are called the *conditions of compatibility*; r and s are also known as *Riemann invariants*. For hydrogen, with $\gamma = \frac{5}{3}$, the conditions are :—

along $$\frac{dx}{dt} = u + c, \quad 3c + u = 2r = \text{constant}, \tag{6a}$$

along $$\frac{dx}{dt} = u - c, \quad 3c - u = 2s = \text{constant}. \tag{6b}$$

The two families of curves will be referred to below as the r-characteristics and s-characteristics respectively. Through any point P in the (x, t) plane at which there is gas there will be two characteristics, one of each family. The characteristics represent the limits of influence of an infinitesimal disturbance originating at P. Given values of u, c at a point $(x, 0)$, the values of r, s corresponding to the characteristics through this point on the line $t = 0$ are known.

At the intersection of an r-characteristic with an s-characteristic the state of the gas is given by the simultaneous solution of the conditions of compatibility in (5 *a*, *b*), viz. :—

$$u = r - s, \quad c = (\gamma - 1)(r + s)/2. \tag{7}$$

If there is a range of x for which the initial conditions are homogeneous, all the r-characteristics stemming from this region have the same value of r, and similarly all the s-characteristics have the same value of s. The equations (5) therefore show that if an r-characteristic (say) from a non-homogeneous region meets successively the s-characteristics from a homogeneous region, the values of u, c along it will be unvarying and hence such a characteristic will be straight. A group of such characteristics constitutes what is termed a *simple wave*. In general, both u and c are varying along a particular characteristic.

If it should happen that two characteristics of the same family intersect one another, the continuity of the motion breaks down. The values of u and c are no longer uniquely determined by the conditions of compatibility. A shock-wave begins at the point and time when such an intersection first occurs, and moves in a determinable way through the gas.*

3. *Burgers' solution.*—Burgers considers a semi-infinite cloud bordering upon a vacuum. The gas is supposed to be initially at rest and to be homogeneous. Thus, at $t = 0$, $u = 0$ and $c = a$ for $x < 0$, if $x = 0$ is the initial front of the cloud and the positive x-axis is directed into the vacuum. Although there is here an initial discontinuity in density at $x = 0$, the discontinuity exists only instantaneously, all values of pressure and density down to the vacuum state being immediately propagated with the appropriate velocity in a simple wave. The solution is illustrated in a diagram (Fig. 1).

A section by $t =$ constant contains homogeneous gas at rest, then a section of a simple wave leading to the vacuum. The wave of expansion is represented by a fan of characteristics centred on O. The s-characteristics through the points $t = 0$, $x < 0$ are straight with gradient $dx/dt = -a$. The r-characteristics through these points are straight until they meet the s-characteristics of the expansion fan. They then curve as shown. The characteristics of the fan itself meet only r-characteristics from an initially homogeneous region; they are therefore straight and constitute a simple wave. At the front of the gas $c = 0$; hence

* A. F. Pillow, *Proc. Camb. Phil. Soc.*, **45**, 558–586, 1949.

$u=2r=2a/(\gamma-1)$. Thus, the particle velocity in the surface of the cloud is constant and equal to $2a/(\gamma-1)$ (i.e., $3a$ for $\gamma=\frac{5}{3}$). This is also the gradient of the leading characteristic. The rarefaction spreads backwards as well as forwards and a glance at Fig. 1 shows that the boundary of moving gas itself moves into the undisturbed gas with a velocity

$$\frac{dx}{dt}=u-c=\frac{3-\gamma}{2}r-\frac{\gamma+1}{2}s=(1-\gamma)r_0=-a$$

(where $r_0=s_0=$ the initial value of r, s on this boundary), i.e., with the speed of sound in the homogeneous cloud. All the r-characteristics which initially set out from the region $x<0$ cut a particular s-characteristic of the fan at the same angle; they eventually touch the leading characteristic at infinity. Thus, small signals from the interior of the cloud, which would be propagated along the characteristics in the (x, t) plane, never quite reach the front of the cloud.

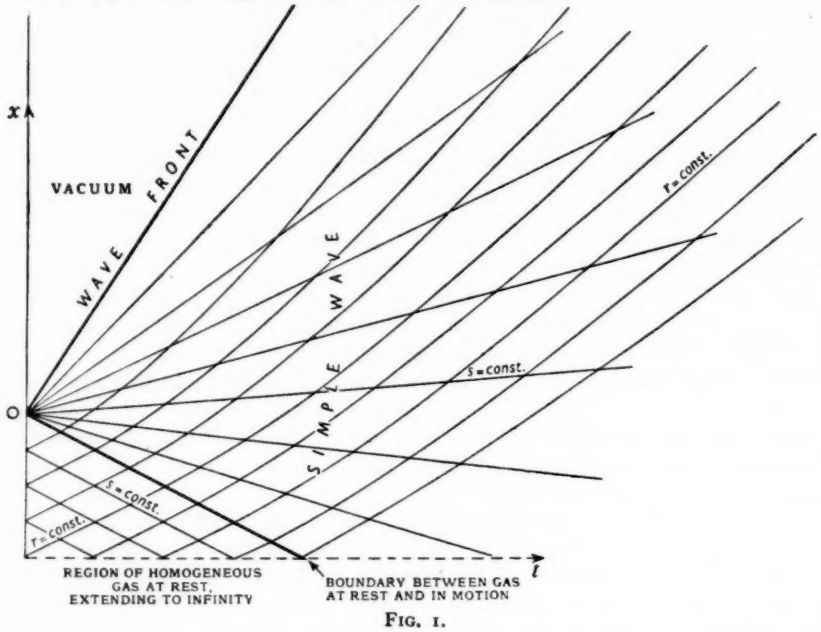

FIG. 1.

The analytical expression for these r-characteristics is given by Courant and Friedrichs*:

$$x=\frac{2}{\gamma-1}at-\frac{\gamma+1}{\gamma-1}at_0(t/t_0)^{2/(\gamma+1)}$$

where $t=t_0$ is the instant at which such a characteristic first meets the fan. For a real gas, since $\gamma>1$, this shows that the whole gas finally has the velocity given by $\lim_{t\to\infty}(dx/dt)_{r=r_0}=2a/(\gamma-1)$. This takes the value $3a$ for $\gamma=\frac{5}{3}$.

4. *Inhomogeneous initial conditions.*—In order to remove the initial discontinuity of velocity, McVittie† suggested the introduction of an initial inhomogeneous

* R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*, New York, 1948.
† *Loc. cit.*

distribution of density in the gas cloud at rest at time $t=0$, but without any discontinuities. It is clear from a characteristics diagram that the inhomogeneous gas must instantly begin to move; for the intersection of r- and s-characteristics of differing constants immediately gives a non-zero value for the particle velocity u.

Copson has shown, in his paper, that the variation of x and t along characteristics in the region in which both r and s vary may be expressed in terms of an auxiliary function ϕ defined by

$$\phi(r)=\int_0^r 2r\,[x(r)]_{t=0}\,dr \quad (r\geqslant 0),\ \phi(-r)=\phi(r), \qquad (\gamma=\tfrac{5}{3})$$

where $[x(r)]_{t=0}$ is the value of x at $t=0$ expressed as a function of the Riemann invariant r. Then

$$\tfrac{3}{4}t=\frac{\partial}{\partial s}\left\{\frac{\phi'(r)}{r+s}+\frac{\phi(s)-\phi(r)}{(r+s)^2}\right\}=\frac{\partial\xi}{\partial s}\ (\text{say})$$

$$=-\frac{\partial}{\partial r}\left\{\frac{\phi'(s)}{r+s}-\frac{\phi(s)-\phi(r)}{(r+s)^2}\right\}=-\frac{\partial\eta}{\partial r}\ (\text{say}),$$

and $x=\xi+(4r-2s)t/3=\eta+(2r-4s)t/3$.

On the face of the cloud $r+s=0$, since the velocity of sound vanishes. A characteristic $r=r_1$ meets the face at a time $t=t_1$ given by

$$\phi'''(r_1)=-4t_1. \tag{8}$$

This formula was given by Copson and follows, by taking a simple limit*, from the expression for t.

Now, if (8) is to give $t_1\geqslant 0$, clearly $\phi'''(r_1)\leqslant 0$. At $t=0$, $x(r)=\phi'(r)/2r$ and $r=c=1/w$ (say). After a little calculation the condition is found to be

$$\left[\frac{\partial^2 x}{\partial w^2}\right]_{t=0}\leqslant 0. \tag{9}$$

If it is supposed that r is a monotonic decreasing function of x then a further necessary condition for preservation of the continuity of the motion is that the time $t=t_1$ increases with r_1; physically this requires that wavelets of the same family do not cross before reaching the surface of the cloud, but arrive at the surface later the farther they were from it at $t=0$. Thus $dt_1/dr_1>0$, implying $\phi^{(\mathrm{IV})}(r_1)<0$, or the equivalent inequality

$$3\frac{\partial^2 x}{\partial w^2}+w\frac{\partial^3 x}{\partial w^3}>0. \tag{10}$$

The conditions (9), (10) are both necessary for continuity of the flow; they are not likely to be sufficient, except perhaps for simple distributions of $[x(r)]_{t=0}$.

5. *Copson's solution.*—Copson discusses the use of Riemann invariants in the solution of the problem for $\gamma=\frac{5}{3}$, and ends with an example of an initial finite inhomogeneity between the vacuum and the semi-infinite extent of homogeneous gas. He supposes that $[r(x)]_{t=0}$ increases parabolically from zero at the front ($x=h$) to the value $b=3a/2$ at $x=0$, and remains constant for $x<0$. In $0\leqslant x\leqslant h$, $r=b\sqrt{(1-x/h)}$. This distribution satisfies the requirements of the previous section and the corresponding solution is represented graphically by a diagram (Fig. 2). The initial inhomogeneity is confined to the strip OA.

* The appropriate ϕ for negative values of s is $\phi(-s)$. This satisfies the requirements of the continuation of the solution across the characteristic $s=0$ and ensures a finite value of t for the arrival of the characteristic $r=r_1$ at the surface.

On the face of the cloud at any time t, $c=0$. An r-characteristic which meets the face (a possibility in this case) reflects at this free boundary into an s-characteristic; since $c=0$, it follows that $s=-r$ and $dx/dt=u$ for both r- and s-characteristics. Hence, (i) the r-characteristics envelop the vacuum face in the (x, t) plane, and (ii) the velocity of advance of the cloud is equal to the local particle velocity ($2r$). The maximum velocity of advance of the front corresponds to the maximum value of r (i.e. $3a/2$); the maximum velocity is therefore $3a$ as before. In the characteristics diagram a section by $t=0$ yields a uniform region, a "mixed" region and a vacuum. In the uniform region both r and s are constant; hence, the boundary OD is straight and has $dx/dt=-a$. Let B be the point at which the r-characteristic from O touches the wave-front. From the discussion of the previous section, a simple

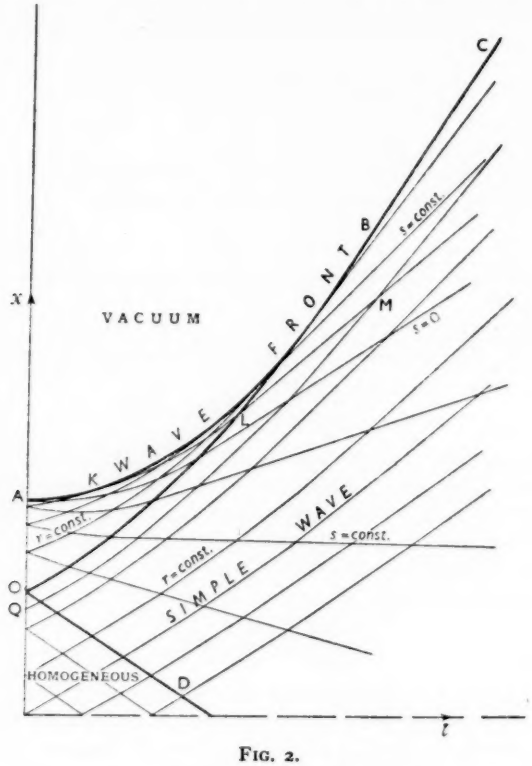

FIG. 2.

wave (but no longer centred on a point) must fill the whole plane to the right of OBC, first of all (in time and space) because of the intersections of r-characteristics from a uniform region with the s-characteristics which come from between O and A; and then because of their intersections with the s-characteristics which are the continuations of r-characteristics after touching the front of the cloud (e.g. KLM).

To prove that the characteristic OBC does, in fact, constitute the boundary from the point B onwards, consider an r-characteristic QM arbitrarily chosen from those to the right of OB. OB and such r-characteristics are all equally inclined to a given s-characteristic; hence QM meets the reflection BC at infinity. QM was,

however, arbitrarily chosen. Hence, OBC continues, after touching the front, as the front itself and BC is straight. From the time at B, t_b, onwards, the mixed region is no longer present; a section at time $t \geqslant t_b$ consists only of a simple wave lying between a uniform region and a vacuum.

An example of an inhomogeneity extending to $x = -\infty$ was given in a paper by Martin* which appears to have been overlooked by earlier writers. He gives the solution for $\gamma = \frac{5}{3}$ and $r = \sqrt{(-x/2)}$ at $t = 0$ $(x \leqslant 0)$ in an ideal gas bordering upon a vacuum. This initial condition satisfies the necessary conditions for continuity; the flow is in fact continuous and the lines of constant density (which, of course, include the surface of the vacuum) are parabolas.

6. *Alternative initial conditions.*—Two further examples of flow with $\gamma = \frac{5}{3}$ are illustrated in Figs. 3 and 4. Fig. 3 corresponds to the case where the speed of sound increases linearly between $x = h$ and $x = 0$. Equation (8) indicates that t_1 is constant for this case, and in fact the r-characteristics of the mixed region now converge to a point B. A point-centred simple wave at B again reduces the pressure at the wave-front to zero. This is the limiting case of Copson's solution when the envelope shrinks to a point. (Burgers' solution is the limiting case when the initial spread of c is confined to a point.)

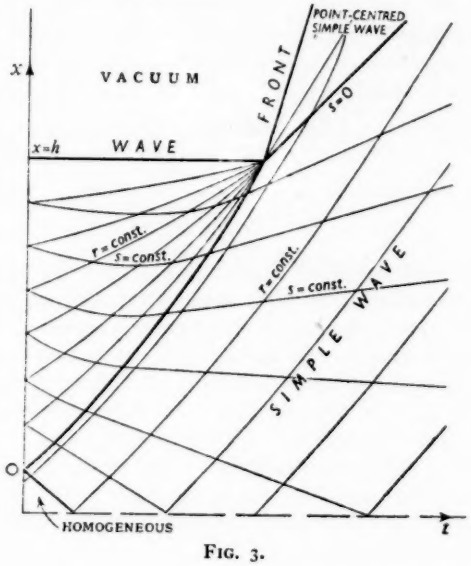

FIG. 3.

Fig. 4 illustrates a further possibility. Here $[x(r)]_{t=0} = (1-r)^2$, $1 \geqslant x \geqslant h$ $(h \geqslant 0)$.† This satisfies the condition (9) only for certain values of r and never satisfies (10); in this case, therefore, the flow cannot remain continuous. The r-characteristics now possess an envelope inside the gas cloud. The intersection of characteristics of the same family gives multiple values of u and c at the points of intersection; this indicates the occurrence of a shock-wave within the flow, beginning at the point and instant when the one-valued character of the flow

* M. H. Martin, *Am. J. Maths.*, **65**, 391–407, 1943.

† The initial point of the shock-wave in Fig. 4 occurs at time $t(h)$ where $t(h)$ moves continuously towards zero as h tends to zero. When $h = 0$ the shock-wave begins instantaneously at $t = 0$.

breaks down. To a first approximation the development of the shock-wave is as shown by the envelope of the characteristics in the diagram.* In an exact treatment the variation of the strength of the shock-wave with time has to be taken into account, the r-characteristics no longer being tangential to the shock-wave after the initial point. A particular case has not been calculated in detail on account of the labour of computation. Certain deductions may, however, be made by a brief examination.

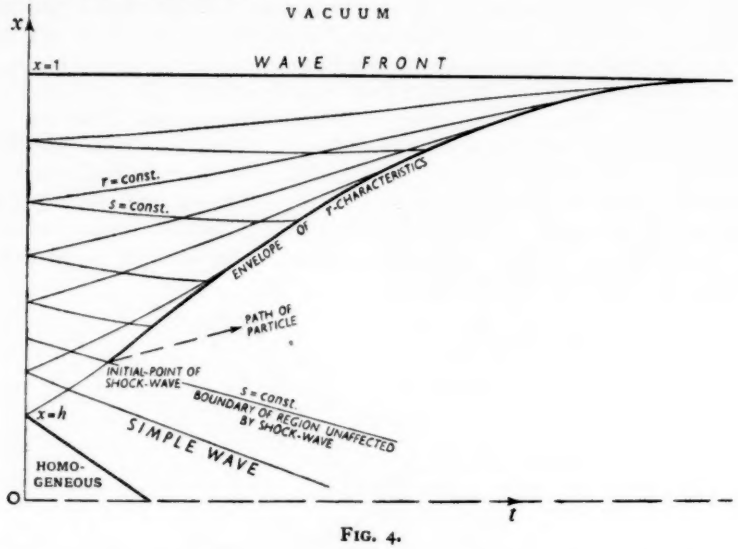

FIG. 4.

As gas crosses the shock-wave it gains in entropy. A dotted line in Fig. 4 represents (diagrammatically) the path of the particle which was at the point of origin of the shock-wave at the critical instant when it began; this line separates those particles which have passed through the shock-wave and experienced an increase of entropy from those which have not. When the shock-wave reaches the front of the cloud a pressure discontinuity exists there; this must, as in Burgers' problem, resolve itself instantly into a wave of expansion. The effect on the velocity of advance of the wave-front may be seen from the condition of compatibility for a characteristic which passes to the right of the shock-wave. Along such a characteristic the relations of Section 2 require modification, and the condition of compatibility is in fact†

$$3c+u-\int_{S_0}^{S} k^2 c\, dS = 3a,$$

where S denotes entropy, S_0 its value in the gas in its original state, and k^2 is a positive constant. Since, by the second law of thermodynamics, S can only increase, and since c is an essentially positive quantity, the value of u when c vanishes is greater than $3a$.

* In the complete mathematical solution the r-characteristics are tangential to the envelope and the s-characteristics are cusped where they meet the envelope. In the diagram the characteristics are only taken up to the envelope; their continuations beyond these points are of no physical significance.

† See, for example, Pillow, *loc. cit.*

7. *Conclusion.*—The motion of a semi-infinite gas cloud bordering upon a vacuum has been examined by the method of characteristics. Characteristics diagrams make physical details of the flow easy to see, and the problems of Burgers and Copson have been examined by means of them. These problems concern a gas with ratio of specific heats $\gamma = \frac{5}{3}$, corresponding to gas clouds of monatomic hydrogen. Burgers' problem involved an initial discontinuity of density at the front of a homogeneous cloud; Copson's stipulated an initial region of inhomogeneity between the vacuum surface and the main homogeneous cloud, without any actual discontinuities in density. Burgers' solution is continuous for $t > 0$; the particular example given by Copson is also continuous. It has been shown above that unless certain conditions are satisfied by the initial state in the cloud the subsequent flow cannot remain indefinitely continuous. An example of the breakdown of continuity is given.

An initially linear distribution of values of the speed of sound in the gas at rest leads to a singular solution in which all the characteristics of one family intersect at a point of the wave-front at the same instant of time $t = T$ (say). The solution is subsequently continuous.

It is shown that, when the flow remains continuous, the wave-front ultimately moves with velocity $3a$. When a shock-wave appears in the flow the wave-front attains an even higher velocity. Thus, the high eventual velocity of advance of these gas clouds, which was considered a physically unsatisfactory feature of the previous solutions, is not to be avoided by variation of the initial inhomogeneities.

*Institute for Fluid Dynamics and Applied Mathematics**,
University of Maryland,
U.S.A.:
1952 *October* 1.

* On leave of absence from University College, Dundee, University of St. Andrews. Now at the University of Manchester.

COHERENT AND NON-COHERENT SCATTERING IN THE THEORY OF LINE FORMATION

Ida W. Busbridge

(Received 1952 October 1)

Summary

When the scattering of light is coherent, the exact solution of the problem of line formation in the Milne–Eddington model, with a linear Planck function $B_\nu = b_\nu^{(0)} + b_\nu^{(1)}\tau_\nu$, has been given by Chandrasekhar. In Section 2 this solution is found by a new method, and in Sections 3–5 the method is used to solve the parallel problem when the scattering of light is completely non-coherent. This solution is exact when $b_\nu^{(0)}$ and $\kappa_\nu b_\nu^{(1)}$ are independent of the frequency ν and the depth z, κ_ν being the continuous absorption coefficient. The solution agrees with that found by Sobolev for the case $\kappa_\nu = \kappa$, $b_\nu^{(0)} = B$, $b_\nu^{(1)} = 0$. In Sections 6–8 the magnitudes of the terms appearing in the solution are discussed. Sobolev's assertion that the resulting central intensities of absorption lines are much lighter than those found on the hypothesis of coherent scattering is shown to be true for lines at the limb. No definite conclusions can be drawn for other lines until the H-functions involved have been computed, but the same result is likely to hold.

---

1. *Introduction.*—In the theory of line formation, it has been assumed, until recently, that the scattering of light is coherent, i.e. that the frequency of the radiation re-emitted within the line depends only on the radiation absorbed at the same frequency. There are, however, several causes leading to a redistribution in frequency, and an investigation of the effects of non-coherent scattering has become necessary. The reasons for this have been fully discussed by G. Münch (**5**) and (more recently) by M. P. Savedoff (**6**). Comprehensive lists of references will be found in both of these papers and Savedoff reviews all previous attempts at solving the problem of line formation when the scattering is either partially or completely non-coherent. Unfortunately he dismisses the important paper by V. V. Sobolev (**7**) as "extremely complicated" and "unclear". Sobolev's paper is certainly "unclear", but it gives the first exact solution of the problem. This is for the case in which scattering is completely non-coherent and the Planck function B_ν and the continuous absorption coefficient κ_ν are both independent of the frequency ν and the depth z.

Savedoff's own method is complicated. He also deals with the case of completely non-coherent scattering and he begins by transforming the equation of transfer (see Section 3) into a non-homogeneous integral equation. He then uses iteration to obtain an approximate solution when the Planck function is linear. The resulting residual intensity in an absorption line gives interesting results for the wings of the line, but appears to become indeterminate in the centre, the place where the discrepancy between theory and observation has always been greatest. This is a weakness in Savedoff's method, though I do not think it should be difficult to overcome the indeterminacy.

In (**7**), Sobolev asserts that the central intensities obtained from his exact solution are a great improvement on those obtained with coherent scattering. For this reason his paper deserves serious consideration.

Sobolev's paper is written in Russian; his assumptions are not clearly stated; the paper contains misprints, and his deductions are questionable, because he neglects terms which are, as a matter of fact (see Section 8), of the same size as those which he retains.

It was whilst trying to elucidate this paper that I began to investigate extensions of Sobolev's results. The case of partially non-coherent scattering can be approximately represented as the sum of coherent and completely non-coherent scatterings in a suitable ratio, and I therefore restricted myself to the case of completely non-coherent scattering, but I replaced Sobolev's assumption that B_ν is a constant by the assumption that B_ν is linear.

The method used by Sobolev breaks down when B_ν is not constant. In order to solve the problem for a linear B_ν, I had first to evolve a new method of solution for the corresponding case of coherent scattering. This is given in Section 2. It uses Chandrasekhar's ideas, but is much simpler than that given in (**2**), § 84, and it can be applied when the scattering is non-coherent. This is done in Sections 3–5.

The solution of the problem (equation (46)) is too complicated to be of much value as it stands. In Sections 6 and 7, I investigate the functions involved in considerable detail. Finally, in Section 8, I consider the sizes of the various terms appearing in (46) and I show that Sobolev's assertion about the improvement in the central intensity is true for lines at the limb. No final conclusion can be drawn until the H-functions involved have been computed.

In an appendix I give an alternative method for finding the scattering function of Section 4. The second method is longer but is easier to follow.

2. *Coherent scattering.*—With the usual notation, the equation of transfer for the Milne–Eddington model can be written

$$\mu \frac{dI_\nu}{\rho\, dz} = (\kappa_\nu + \sigma_\nu) I_\nu - \tfrac{1}{2}\sigma_\nu \int_{-1}^{1} I_\nu\, d\mu - \kappa_\nu B_\nu, \tag{1}$$

where z is the depth below the surface, κ_ν is the continuous absorption coefficient and σ_ν is the line-scattering coefficient. We assume that κ_ν and σ_ν are independent of depth and we write

$$t_\nu = \int_0^z \rho(\kappa_\nu + \sigma_\nu)\, dz, \qquad \tau_\nu = \int_0^z \rho\kappa_\nu\, dz, \tag{2}$$

$$\eta_\nu = \sigma_\nu/\kappa_\nu, \qquad \lambda_\nu = 1/(1+\eta_\nu) = \kappa_\nu/(\kappa_\nu + \sigma_\nu). \tag{3}$$

Then $\tau_\nu = \lambda_\nu t_\nu$ and

$$B_\nu = b_\nu^{(0)} + b_\nu^{(1)}\tau_\nu = b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)} t_\nu, \tag{4}$$

if we take the first two terms in the Taylor expansion of B_ν. Substituting into (1) we get *

$$\mu \frac{dI_\nu(t_\nu, \mu)}{dt_\nu} = I_\nu(t_\nu, \mu) - \tfrac{1}{2}(1-\lambda_\nu) \int_{-1}^{1} I_\nu(t_\nu, \mu')\, d\mu' - \lambda_\nu(b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)} t_\nu). \tag{5}$$

* This is (**2**), § 84, equation (24), except that Chandrasekhar drops the subscript ν and allows for thermal emission. We shall retain ν because it becomes a variable in Sections 3–8. The result of allowing for thermal emission is to change the definition (3) of λ_ν into $\lambda_\nu = (1+\epsilon_\nu\eta_\nu)/(1+\eta_\nu)$. Then $\tau_\nu = \lambda_\nu t_\nu/(1+\epsilon_\nu\eta_\nu)$ and $b_\nu^{(1)}$ has to be replaced by $b_\nu^{(1)}/(1+\epsilon_\nu\eta_\nu)$. On making these changes in the final solution (16), the required solution is obtained.

The same changes apply to the equation of transfer (19) for non-coherent scattering, but because this still contains σ_ν', the changes cannot be immediately applied to the solution. It is not difficult, however, to introduce the necessary modifications into the subsequent analysis. We shall take $\epsilon_\nu = 0$ throughout.

The equation (5) has to be solved subject to the boundary condition

$$I_\nu(0, -\mu') \equiv 0 \qquad (0 < \mu' \leqslant 1) \tag{6}$$

together with the condition (implicit in Chandrasekhar's solution) that $I_\nu(t_\nu, \mu)$ is at most linear in t_ν for large t_ν. Now a particular solution of (5), which does not satisfy (6), is

$$I_\nu(t_\nu, \mu) = b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)}(t_\nu + \mu), \tag{7}$$

as is readily verified by substitution. We therefore write

$$I_\nu(t_\nu, \mu) = b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)}(t_\nu + \mu) + I_\nu^*(t_\nu, \mu). \tag{8}$$

Then $I_\nu^*(t_\nu, \mu)$ satisfies the integro-differential equation

$$\mu \frac{dI_\nu^*(t_\nu, \mu)}{dt_\nu} = I_\nu^*(t_\nu, \mu) - \tfrac{1}{2}(1 - \lambda_\nu) \int_{-1}^{1} I_\nu^*(t_\nu, \mu')\, d\mu', \tag{9}$$

together with the boundary condition

$$I_\nu^*(0, -\mu') \equiv \lambda_\nu b_\nu^{(1)} \mu' - b_\nu^{(0)} \qquad (0 < \mu' \leqslant 1). \tag{10}$$

Moreover $I_\nu^*(t_\nu, \mu)$ must be at most linear in t_ν for large t_ν.

We require the emergent intensity $I_\nu^*(0, \mu)$. This is the sum of $I_1^*(0, \mu)$, where $I_1^*(t_\nu, \mu)$ is the solution of (9) subject to the boundary condition

$$I_1^*(0, -\mu') \equiv 0 \qquad (0 < \mu' \leqslant 1), \tag{11}$$

and $I_2^*(0, \mu)$, where this is the diffusely reflected intensity corresponding to the incident intensity (10).

It can be shown that †, unless $\lambda_\nu = 0$ (which is not so),

$$I_1^*(t_\nu, \mu) \equiv 0.$$

Hence

$$\begin{aligned} I_\nu^*(0, \mu) &= I_2^*(0, \mu) \\ &= \frac{1}{2\mu} \int_0^1 S(\mu, \mu')(\lambda_\nu b_\nu^{(1)} \mu' - b_\nu^{(0)})\, d\mu', \end{aligned} \tag{12}$$

where by (**2**), § 33, equation (41),

$$S(\mu, \mu') = (1 - \lambda_\nu) \frac{\mu\mu'}{\mu + \mu'} H(\mu) H(\mu') \tag{13}$$

and $H(\mu)$ is the solution of

$$H(\mu) = 1 + \tfrac{1}{2}(1 - \lambda_\nu)\mu H(\mu) \int_0^1 \frac{H(\mu')}{\mu + \mu'}\, d\mu'. \tag{14}$$

From (12) and (13) we have

$$\begin{aligned} I_\nu^*(0, \mu) &= \tfrac{1}{2}(1 - \lambda_\nu) H(\mu) \int_0^1 (\lambda_\nu b_\nu^{(1)} \mu' - b_\nu^{(0)}) \frac{\mu' H(\mu')}{\mu + \mu'}\, d\mu' \\ &= \tfrac{1}{2}(1 - \lambda_\nu)\lambda_\nu b_\nu^{(1)} H(\mu) \int_0^1 \left(\mu' - \mu + \frac{\mu^2}{\mu + \mu'} \right) H(\mu')\, d\mu' \\ &\quad - \tfrac{1}{2}(1 - \lambda_\nu) b_\nu^{(0)} H(\mu) \int_0^1 \left(1 - \frac{\mu}{\mu + \mu'} \right) H(\mu')\, d\mu', \end{aligned}$$

† This can be seen as follows: with the notation of (**2**), § 26,

$$I_1^*(t_\nu, \mu) \simeq \tfrac{1}{2}(1 - \lambda_\nu) \sum_{\alpha=1}^{n} L_\alpha e^{-k_\alpha t_\nu} / (1 + \mu k_\alpha),$$

where the L_α are determined by

$$\sum_{\alpha=1}^{n} L_\alpha / (1 - \mu_i k_\alpha) = 0 \qquad (i = 1, \ldots, n).$$

Since $\prod_{\alpha=1}^{n} (1 - \mu k_\alpha) \times \sum_{\alpha=1}^{n} L_\alpha / (1 - \mu k_\alpha)$ is a polynomial in μ of degree $n - 1$ with n distinct zeros, it is identically zero. Hence every $L_\alpha = 0$, and in the limit, as $n \to \infty$, $I_1^*(t_\nu, \mu) = 0$.

and by (14) this is

$$I_\nu^*(0,\mu)=(b_\nu^{(0)}+\lambda_\nu b_\nu^{(1)}\mu)\{H(\mu)-1-\tfrac{1}{2}(1-\lambda_\nu)\alpha_0 H(\mu)\}+\tfrac{1}{2}(1-\lambda_\nu)\lambda_\nu b_\nu^{(1)}\alpha_1 H(\mu), \quad (15)$$

where

$$\alpha_n=\int_0^1 \mu^n H(\mu)\,d\mu.$$

By (**2**), §38, Theorem 1,

$$1-\tfrac{1}{2}(1-\lambda_\nu)\alpha_0=\lambda_\nu^{1/2},$$

and so (8) and (15) give

$$I_\nu(0,\mu)=(b_\nu^{(0)}+\lambda_\nu b_\nu^{(1)}\mu)\lambda_\nu^{1/2}H(\mu)+\tfrac{1}{2}(1-\lambda_\nu)\lambda_\nu b_\nu^{(1)}\alpha_1 H(\mu). \quad (16)$$

This is Chandrasekhar's solution ((**2**), §84.4, equation (66)).

3. *The equation of transfer in the case of completely non-coherent scattering.*— When the scattering is completely non-coherent, the scattering term on the right-hand side of (1) is replaced by

$$\tfrac{1}{2}\sigma_\nu\int_0^\infty \sigma_{\nu'}\,d\nu'\int_{-1}^{1} I_{\nu'}\,d\mu'\Big/\int_0^\infty \sigma_{\nu'}\,d\nu',$$

and the equation of transfer is

$$\mu\frac{dI_\nu}{\rho\,dz}=(\kappa_\nu+\sigma_\nu)I_\nu-\tfrac{1}{2}\sigma_\nu\int_0^\infty \sigma_{\nu'}\,d\nu'\int_{-1}^{1} I_{\nu'}\,d\mu'\Big/\int_0^\infty \sigma_{\nu'}\,d\nu'-\kappa_\nu B_\nu. \quad (17)$$

In this we can, without loss of generality, take

$$\int_0^\infty \sigma_{\nu'}\,d\nu'=1. \quad (18)$$

In doing this we are really replacing σ_ν, κ_ν and ρ by

$$\sigma_\nu\Big/\int_0^\infty \sigma_{\nu'}\,d\nu', \qquad \kappa_\nu\Big/\int_0^\infty \sigma_{\nu'}\,d\nu', \qquad \rho\int_0^\infty \sigma_{\nu'}\,d\nu',$$

but t_ν, τ_ν, η_ν and λ_ν are unchanged.

On using (2), (3), (4) and (18), (17) becomes

$$\mu\frac{dI_\nu(t_\nu,\mu)}{dt_\nu}=I_\nu(t_\nu,\mu)-\tfrac{1}{2}(1-\lambda_\nu)\int_0^\infty \sigma_{\nu'}\,d\nu'\int_{-1}^{1} I_{\nu'}(t_{\nu'},\mu')\,d\mu'-\lambda_\nu(b_\nu^{(0)}+\lambda_\nu b_\nu^{(1)}t_\nu). \quad (19)$$

This has to be solved subject to the boundary condition (6) and the condition that $I_\nu(t_\nu,\mu)$ is at most linear in t_ν for large t_ν.

In order to find $I_\nu(0,\mu)$ by the method of Section 2, we have to find a particular solution corresponding to (7) and a scattering function corresponding to (13). We shall deal first with the latter.

4. *The scattering function for completely non-coherent scattering.*—In this section I shall follow the method due to Ambartsumian (**1**) which was used by Sobolev. This method is explained in great detail in (**3**), §31, and I have introduced modifications in Sobolev's analysis in order to bring the notation into line with that of (**2**) and (**3**). An alternative method following that used by Chandrasekhar in (**2**) is given in the Appendix.

In Sobolev's paper (**7**), κ_ν is assumed to be independent of ν but this is an unnecessary assumption.

Consider a semi-infinite, plane-parallel, scattering atmosphere subject to incident radiation of frequency ν_0 all of whose rays make a fixed angle $\cos^{-1}\mu_0$ ($0<\mu_0\leqslant 1$) with the inwards-drawn normal. Let the monochromatic flux per unit area perpendicular to the rays be F_{ν_0}. Then the incident intensity for any direction $-\mu'$ and frequency ν' is represented exactly by

$$I_\nu(0,-\mu')=\tfrac{1}{2}F_{\nu_0}\delta(\nu_0-\nu')\delta(\mu_0-\mu'), \quad (20)$$

where δ is Dirac's δ-function.

Let the scattering function for the atmosphere be

$$S(\nu, \nu', \mu, \mu'),$$

so that the diffusely reflected intensity $I_\nu^{\text{ref}}(0, \mu)$ of frequency ν and direction μ, due to an incident intensity $I_{\nu'}^{\text{inc}}(0, -\mu')$ of frequency ν' and direction $-\mu'$, is given by

$$I_\nu^{\text{ref}}(0, \mu) = \int_0^\infty d\nu' \int_0^1 I_{\nu'}^{\text{inc}}(0, -\mu') \frac{S(\nu, \nu', \mu, \mu')}{2\mu} d\mu'. \tag{21}$$

In particular the diffusely reflected intensity due to the incident intensity (20) is

$$F_{\nu_0} S(\nu, \nu_0, \mu, \mu_0)/4\mu.$$

Consider the passage of the radiation through the layer of thickness dz below the surface of the atmosphere. The equation of transfer is (17) without the term $\kappa_\nu B_\nu$. This can be written

$$dI_\nu = \frac{\rho\, dz}{\mu}(\kappa_\nu + \sigma_\nu) I_\nu - \frac{\rho\, dz}{2\mu} \sigma_\nu \int_0^\infty \sigma_{\nu'} d\nu' \int_{-1}^1 I_{\nu'} d\mu',$$

from which the absorption and scattering processes can be seen. By the principle of invariance, radiation travelling downwards through the layer is diffusely reflected in the lower surface of the layer according to equation (21). After tracing all processes in the passage of the radiation, we get the following equation (to the first order in dz) for $S(\nu, \nu_0, \mu, \mu_0)$:

$$\begin{aligned}\frac{S(\nu, \nu_0, \mu, \mu_0)}{4\mu} &= \frac{S(\nu, \nu_0, \mu, \mu_0)}{4\mu}\left[1 - (\sigma_{\nu_0} + \kappa_{\nu_0})\frac{\rho\, dz}{\mu_0}\right]\left[1 - (\sigma_\nu + \kappa_\nu)\frac{\rho\, dz}{\mu}\right] + \frac{\sigma_\nu \sigma_{\nu_0} \rho\, dz}{4\mu} \\ &\quad + \frac{\sigma_{\nu_0} \rho\, dz}{4} \int_0^\infty \sigma_{\nu'} d\nu' \int_0^1 \frac{S(\nu, \nu', \mu, \mu')}{2\mu} \frac{d\mu'}{\mu'} + \frac{\sigma_\nu \rho\, dz}{4\mu} \int_0^\infty \sigma_{\nu'} d\nu' \\ &\quad \times \int_0^1 \frac{S(\nu', \nu_0, \mu', \mu_0)}{2\mu'} d\mu' + \frac{\rho\, dz}{4} \int_0^\infty \sigma_{\nu''} d\nu'' \int_0^1 \frac{S(\nu, \nu'', \mu, \mu'')}{2\mu} \frac{d\mu''}{\mu''} \\ &\quad \times \int_0^\infty \sigma_{\nu'} d\nu' \int_0^1 \frac{S(\nu', \nu_0, \mu', \mu_0)}{2\mu'} d\mu'.\end{aligned}$$

In the limit, as $dz \to 0$, this gives

$$\begin{aligned}\frac{S(\nu, \nu_0, \mu, \mu_0)}{\sigma_\nu \sigma_{\nu_0}}\left\{\frac{\sigma_\nu + \kappa_\nu}{\mu} + \frac{\sigma_{\nu_0} + \kappa_{\nu_0}}{\mu_0}\right\} &= \left(1 + \frac{1}{2}\int_0^\infty \sigma_{\nu'}^2\, d\nu' \int_0^1 \frac{S(\nu, \nu', \mu, \mu')}{\sigma_\nu \sigma_{\nu'}} \frac{d\mu'}{\mu'}\right) \\ &\quad \times \left(1 + \frac{1}{2}\int_0^\infty \sigma_{\nu'}^2\, d\nu' \int_0^1 \frac{S(\nu', \nu_0, \mu', \mu_0)}{\sigma_{\nu'} \sigma_{\nu_0}} \frac{d\mu'}{\mu'}\right).\end{aligned} \tag{22}$$

By the principle of reciprocity,

$$S(\nu, \nu, \mu_0, \mu) = S(\nu, \nu, \mu, \mu_0).$$

From this and (22), by a rather lengthy argument which I omit, it can be shown that

$$S(\nu_0, \nu, \mu_0, \mu) = S(\nu, \nu_0, \mu, \mu_0),$$

and so, if

$$\phi(\nu, \mu) = 1 + \frac{1}{2}\int_0^\infty \sigma_{\nu'}^2\, d\nu' \int_0^1 \frac{S(\nu, \nu', \mu, \mu')}{\sigma_\nu \sigma_{\nu'}} \frac{d\mu'}{\mu'}, \tag{23}$$

then (22) gives

$$\frac{S(\nu, \nu_0, \mu, \mu_0)}{\sigma_\nu \sigma_{\nu_0}} = \frac{\mu\mu_0 \phi(\nu, \mu)\phi(\nu_0, \mu_0)}{(\sigma_\nu + \kappa_\nu)\mu_0 + (\sigma_{\nu_0} + \kappa_{\nu_0})\mu}. \tag{24}$$

On substituting from (24) into (23), we have

$$\phi(\nu,\mu) = 1 + \tfrac{1}{2}\mu\phi(\nu,\mu)\int_0^\infty \sigma_{\nu'}^2\, d\nu' \int_0^1 \frac{\phi(\nu',\mu')}{(\sigma_\nu+\kappa_\nu)\mu' + (\sigma_{\nu'}+\kappa_{\nu'})\mu}\, d\mu'. \tag{25}$$

From this it follows that $\phi(\nu,\mu)$ is a function of $\mu/(\sigma_\nu+\kappa_\nu)$.

Let

$$\kappa = \min_\nu (\kappa_\nu + \sigma_\nu) \tag{26}$$

and write

$$x = \frac{\kappa\mu}{\kappa_\nu+\sigma_\nu}, \qquad x' = \frac{\kappa\mu'}{\kappa_{\nu'}+\sigma_{\nu'}}, \tag{27}$$

$$H(x) = \phi(\nu,\mu). \tag{28}$$

Then (25) becomes

$$H(x) = 1 + \frac{x}{2\kappa} H(x) \int_0^\infty \sigma_{\nu'}^2\, d\nu' \int_0^{\kappa/(\sigma_{\nu'}+\kappa_{\nu'})} \frac{H(x')}{x+x'}\, dx'. \tag{29}$$

When the order of integration is changed in (29), the limits for ν' have to be changed to ones depending on the form of the curve

$$x' = \kappa/(\kappa_{\nu'}+\sigma_{\nu'}) \qquad (\kappa = \min_\nu (\kappa_\nu+\sigma_\nu)).$$

The line scattering coefficient $\sigma_{\nu'}$ is symmetrical about the centre $\nu' = \nu_c$ of the line considered and it decreases rapidly to a very small value on either side. No appreciable error is therefore introduced if (29) is inverted into

$$H(x) = 1 + \tfrac{1}{2} x H(x) \int_0^1 \frac{H(x')K(x')}{x+x'}\, dx', \tag{30}$$

where

$$K(x') = \frac{2}{\kappa}\int_{\nu(x')}^\infty \sigma_{\nu'}^2\, d\nu' \tag{31}$$

and

$$\left.\begin{aligned} \nu(x') &= \nu_c && \text{if} \quad x' < \kappa/(\kappa_{\nu_c}+\sigma_{\nu_c}), \\ x' &= \kappa/(\kappa_{\nu(x')}+\sigma_{\nu(x')}) && \text{if} \quad x' > \kappa/(\kappa_{\nu_c}+\sigma_{\nu_c}). \end{aligned}\right\} \tag{32}$$

In terms of the new variables, the scattering function (24) is given by

$$S(\nu,\nu',\mu,\mu') = \frac{\sigma_\nu\sigma_{\nu'}}{\kappa} \cdot \frac{xx'H(x)H(x')}{x+x'}, \tag{33}$$

where x and x' are given by (27).

If κ_ν is a constant for all ν, then $\kappa_\nu = \kappa$ since $\sigma_\nu \to 0$ as $|\nu-\nu_c| \to \infty$. The formulae (30)–(32) are then identical with those of Sobolev.

When the function $K(x')$ has been determined, the solution $H(x)$ of (30) can be found by the formulae given in (**2**), Chap. V. Note that the $H(x)$ of (30) is not the same function as the $H(\mu)$ of (14).

5. *The solution of equation* (19).—If, following the method of Section 2, we look for a particular solution of (19) similar to (7) by substituting

$$I_\nu(t_\nu,\mu) = A_\nu + B_\nu t_\nu + C_\nu\mu,$$

where A_ν, B_ν, C_ν are independent of t_ν and μ, we find that there is little hope of succeeding, even with more elaborate functions, except in special cases, and then (7) is itself a solution. Let us therefore write, as in Section 2,

$$I_\nu(t_\nu,\mu) = b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)}(t_\nu+\mu) + I_\nu^*(t_\nu,\mu). \tag{34}$$

Then on using (18) it is found that $I_\nu^*(t_\nu, \mu)$ must satisfy the equation

$$\mu \frac{dI_\nu^*(t_\nu, \mu)}{dt_\nu} = I_\nu^*(t_\nu, \mu) - \tfrac{1}{2}(1-\lambda_\nu) \int_0^\infty \sigma_{\nu'}\, d\nu' \int_{-1}^{1} I_{\nu'}^*(t_{\nu'}, \mu')\, d\mu' + g(t_\nu, \mu), \quad (35)$$

where

$$g(t_\nu, \mu) = (1-\lambda_\nu) \int_0^\infty \sigma_{\nu'} \{(b_\nu^{(0)} - b_{\nu'}^{(0)}) + (\lambda_\nu b_\nu^{(1)} t_\nu - \lambda_{\nu'} b_{\nu'}^{(1)} t_{\nu'})\}\, d\nu'$$

$$= (1-\lambda_\nu) \int_0^\infty \sigma_{\nu'} \{(b_\nu^{(0)} - b_{\nu'}^{(0)}) + (b_\nu^{(1)} \kappa_\nu - b_{\nu'}^{(1)} \kappa_{\nu'}) \int_0^z \rho\, dz\}\, d\nu'. \quad (36)$$

Now $\sigma_{\nu'}$ is greatest at $\nu' = \nu_c$ and decreases rapidly to a very small value on either side. If ν is near to ν_c, $b_\nu^{(0)} - b_{\nu'}^{(0)}$ and $b_\nu^{(1)} \kappa_\nu - b_{\nu'}^{(1)} \kappa_{\nu'}$ are both small when $\sigma_{\nu'}$ is not small, and the integrand in (36) is small over the entire range of integration. This suggests that $g(t_\nu, \mu)$ is negligible for values of ν near to ν_c and those are the important values in the theory of spectral lines.

The function $g(t_\nu, \mu)$ is identically zero if $b_\nu^{(0)}$ and $b_\nu^{(1)} \kappa_\nu$ are independent of ν, and then the following equations are exact. Sobolev's results (but not his method) are obtained by taking $b_\nu^{(0)} = B$ (a constant), $b_\nu^{(1)} = 0$.

If we neglect $g(t_\nu, \mu)$, we have

$$\mu \frac{dI_\nu^*(t_\nu, \mu)}{dt_\nu} = I_\nu^*(t_\nu, \mu) - \tfrac{1}{2}(1-\lambda_\nu) \int_0^\infty \sigma_{\nu'}\, d\nu' \int_{-1}^{1} I_{\nu'}^*(t_{\nu'}, \mu')\, d\mu', \quad (37)$$

together with the boundary condition

$$I_{\nu'}^*(0, -\mu') \equiv \lambda_{\nu'} b_{\nu'}^{(1)} \mu' - b_{\nu'}^{(0)} \qquad (0 < \mu' \leqslant 1, \quad 0 < \nu' < \infty) \quad (38)$$

and the condition that $I_\nu^*(t_\nu, \mu)$ must be at most linear for large t_ν. Now, as in Section 2, the only solution $I_1^*(t_\nu, \mu)$ of (37) subject to the boundary condition

$$I_1^*(0, -\mu') \equiv 0 \qquad (0 < \mu' \leqslant 1, \quad 0 < \nu' < \infty)$$

is $I_1^*(t_\nu, \mu) \equiv 0$. Hence the required solution of (37) is the diffusely reflected intensity due to the incident intensity (38), i.e. by (21)

$$I_\nu^*(0, \mu) = \int_0^\infty d\nu' \int_0^1 (\lambda_{\nu'} b_{\nu'}^{(1)} \mu' - b_{\nu'}^{(0)}) \frac{S(\nu, \nu', \mu, \mu')}{2\mu}\, d\mu'.$$

By (33) and (27), this is

$$I_\nu^*(0, \mu) = \tfrac{1}{2} H(x) \frac{\sigma_\nu}{\sigma_\nu + \kappa_\nu} \int_0^\infty \sigma_{\nu'}\, d\nu' \int_0^{\kappa/(\kappa_{\nu'} + \sigma_{\nu'})} \frac{x' H(x')}{x + x'}$$

$$\times \left[\lambda_{\nu'} b_{\nu'}^{(1)} (\kappa_{\nu'} + \sigma_{\nu'}) \frac{x'}{\kappa} - b_{\nu'}^{(0)} \right] (\kappa_{\nu'} + \sigma_{\nu'}) \frac{dx'}{\kappa}$$

$$= \tfrac{1}{2} H(x) \frac{1 - \lambda_\nu}{\kappa} \int_0^\infty (\kappa_{\nu'} + \sigma_{\nu'}) \sigma_{\nu'}\, d\nu' \int_0^{\kappa/(\kappa_{\nu'} + \sigma_{\nu'})} \frac{x' H(x')}{x + x'} \left(b_{\nu'}^{(1)} \kappa_{\nu'} \frac{x'}{\kappa} - b_{\nu'}^{(0)} \right) dx'$$

by (3).

Since $\sigma_{\nu'}$ is large only near $\nu' = \nu_c$, we can again replace $b_{\nu'}^{(1)} \kappa_{\nu'}$ and $b_{\nu'}^{(0)}$ by $b_\nu^{(1)} \kappa_\nu$ and $b_\nu^{(0)}$ when ν is near to ν_c, and we have approximately

$$I_\nu^*(0, \mu) = \tfrac{1}{2}(1-\lambda_\nu) H(x) \frac{b_\nu^{(1)} \kappa_\nu}{\kappa^2} \int_0^\infty (\kappa_{\nu'} + \sigma_{\nu'}) \sigma_{\nu'}\, d\nu' \int_0^{\kappa/(\kappa_{\nu'} + \sigma_{\nu'})} \frac{x'^2 H(x')}{x + x'}\, dx'$$

$$- \tfrac{1}{2}(1-\lambda_\nu) H(x) \frac{b_\nu^{(0)}}{\kappa} \int_0^\infty (\kappa_{\nu'} + \sigma_{\nu'}) \sigma_{\nu'}\, d\nu' \int_0^{\kappa/(\kappa_{\nu'} + \sigma_{\nu'})} \frac{x' H(x')}{x + x'}\, dx'. \quad (39)$$

Let *

$$K_1(x') = \frac{2}{\kappa}\int_{\nu(x')}^{\infty} \kappa_\nu \sigma_\nu \, d\nu', \tag{40}$$

where $\nu(x')$ is given by (32). Then inverting the order of integration in (39), we have, by (31) and (40),

$$I_\nu^*(0,\mu) = \tfrac{1}{2}(1-\lambda_\nu)\frac{b_\nu^{(1)}\kappa_\nu}{\kappa} H(x)\int_0^1 \frac{x'^2 H(x')}{x+x'}\{K(x')+K_1(x')\}\,dx'$$
$$-\tfrac{1}{2}(1-\lambda_\nu)b_\nu^{(0)}H(x)\int_0^1 \frac{x'H(x')}{x+x'}\{K(x')+K_1(x')\}\,dx'. \tag{41}$$

This equation is exact when $b_\nu^{(0)}$ and $b_\nu^{(1)}\kappa_\nu$ are independent of ν.

The function $K_1(x')$ is related to $K(x')$ by the equation

$$\int_0^1 K_1(x')\,dx' + \int_0^1 K(x')\,dx' = 1; \tag{42}$$

for

$$\int_0^1 K(x')\,dx' = \frac{1}{\kappa}\int_0^\infty \sigma_\nu^2\,d\nu' \int_0^{\kappa/(\kappa_\nu+\sigma_\nu)} dx' = \int_0^\infty \frac{\sigma_\nu^2}{\kappa_\nu+\sigma_\nu}\,d\nu',$$
$$\int_0^1 K_1(x')\,dx' = \frac{1}{\kappa}\int_0^\infty \sigma_\nu\kappa_\nu\,d\nu' \int_0^{\kappa/(\kappa_\nu+\sigma_\nu)} dx' = \int_0^\infty \frac{\sigma_\nu\kappa_\nu}{\kappa_\nu+\sigma_\nu}\,d\nu', \tag{43}$$

and adding we have (42) by (18). By (3), (43) can be written

$$\int_0^1 K_1(x')\,dx' = \int_0^\infty \lambda_\nu\sigma_\nu\,d\nu'. \tag{44}$$

Since, by (**2**), Chap. V, Theorem 1,

$$\frac{1}{2}\int_0^1 H(x')K(x')\,dx' = 1 - \left[1 - \int_0^1 K(x')\,dx'\right]^{1/2},$$

or, by (42),

$$\frac{1}{2}\int_0^1 H(x')K(x')\,dx' = 1 - \left(\int_0^1 K_1(x')\,dx'\right)^{1/2}, \tag{45}$$

the terms in (41) containing $K(x')$ can be simplified by the process used in obtaining (15). On substituting the resulting expression into (34) we get

$$I_\nu(0,\mu) = \left(b_\nu^{(0)} + \frac{b_\nu^{(1)}\kappa_\nu}{\kappa}x\right)\left\{(1-\lambda_\nu)H(x)\left(\int_0^1 K_1(x')\,dx'\right)^{1/2} + \lambda_\nu\right\}$$
$$+\tfrac{1}{2}(1-\lambda_\nu)H(x)\left\{\frac{b_\nu^{(1)}\kappa_\nu}{\kappa}\int_0^1 x'H(x')K(x')\,dx' + \frac{b_\nu^{(1)}\kappa_\nu}{\kappa}\int_0^1 \frac{x'^2H(x')K_1(x')}{x+x'}\,dx'\right.$$
$$\left. - b_\nu^{(0)}\int_0^1 \frac{x'H(x')K_1(x')}{x+x'}\,dx'\right\}. \tag{46}$$

This is the exact solution when $b_\nu^{(0)}$ and $\kappa_\nu b_\nu^{(1)}$ are independent of ν.

Since $x = \kappa\mu/(\kappa_\nu+\sigma_\nu) = \kappa\mu\lambda_\nu/\kappa_\nu$, therefore

$$b_\nu^{(0)} + \frac{b_\nu^{(1)}\kappa_\nu}{\kappa}x = b_\nu^{(0)} + \lambda_\nu b_\nu^{(1)}\mu.$$

This shows that there is a similarity of form between (46) and (16).

We must now investigate the behaviour of the functions $K(x)$ and $K_1(x)$ and the sizes of the various terms appearing in (46). *In all that follows we shall assume that* $\kappa_\nu = \kappa$.

* A very small error is introduced when $K_1(x')$ is written in this form, just as a small error is introduced when $K(x')$ is written in the form (31). These errors are removed whenever the integrals are re-inverted (as in the proof of (42)).

6. *The functions $K(x)$ and $K_1(x)$.*—When $\kappa_\nu = \kappa$ we have, from (31), (32) and (40),

$$K(x) = \frac{2}{\kappa}\int_{\nu(x)}^{\infty} \sigma_\nu^2\, d\nu, \tag{47}$$

$$K_1(x) = 2\int_{\nu(x)}^{\infty} \sigma_\nu\, d\nu, \tag{48}$$

where

$$\left.\begin{aligned} \nu(x) &= \nu_c \quad \text{if} \quad x < 1/(1+\eta_{\nu_c}),\\ x &= (1+\eta_{\nu(x)})^{-1} \quad \text{if} \quad x > 1/(1+\eta_{\nu_c}), \end{aligned}\right\} \tag{49}$$

and $\eta_{\nu_c} = \sigma_{\nu_c}/\kappa$. This will be large; Sobolev, in an example, takes $\eta_{\nu_c} = 10^6$.

In the notation of (**4**),

$$\sigma_\nu = \sigma_{\nu_c} \cdot \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{-y^2}}{a^2+(\omega-y)^2}\, dy, \tag{50}$$

where

$$\omega = 2(\nu-\nu_c)(\ln 2)^{1/2}/\Delta\nu_D, \qquad a = \Delta\nu_N(\ln 2)^{1/2}/\Delta\nu_D. \tag{51}$$

Since (18) must be true,

$$1 = \int_0^{\infty} \sigma_\nu\, d\nu \simeq \int_{-\infty}^{\infty} \sigma_\nu\, d\nu = \frac{\sigma_{\nu_c} a}{\pi} \cdot \frac{\Delta\nu_D}{2(\ln 2)^{1/2}} \int_{-\infty}^{\infty} e^{-y^2} dy \int_{-\infty}^{\infty} \frac{d\omega}{a^2+(\omega-y)^2}$$

and this at once gives

$$2(\ln 2)^{1/2}/\Delta\nu_D = \pi^{1/2}\sigma_{\nu_c}.$$

From (51) we now have

$$\omega = \pi^{1/2}\sigma_{\nu_c}(\nu-\nu_c). \tag{52}$$

The classical values for a in the visible part of the spectrum are small; values given in (**4**) range from 0·0012 to 0·035. However, P. ten Bruggencate and J. Houtgast have shown (**8**) that the observed value of the natural damping ratio may be as much as thirty times the classical value and so a may be as large as 1·05. The following discussion only applies when a is not greater than 0·01, since then (see the appendix of (**4**)) σ_ν is given approximately by

$$\left.\begin{aligned} \sigma_\nu &= \sigma_{\nu_c} e^{-\omega^2} \quad \text{for} \quad |\omega| \leqslant \omega_0,\\ \sigma_\nu &= \sigma_{\nu_c} a\pi^{-1/2}\omega^{-2} \quad \text{for} \quad |\omega| > \omega_0, \end{aligned}\right\} \tag{53}$$

where, for continuity,

$$e^{-\omega_0^2} = a\pi^{-1/2}\omega_0^{-2}. \tag{54}$$

This equation gives two values for ω_0^2 and we must choose the larger. It is impossible to find an approximate formula for this root and it has therefore to be evaluated for each value of a.

In the following discussion I shall estimate the magnitudes of terms when

$$a = 0{\cdot}01, \qquad \eta_{\nu_c} = 10^6. \tag{55}$$

This is the example chosen by Sobolev. In this case

$$\omega_0 \simeq 2{\cdot}67. \tag{56}$$

First consider $K(x)$. Let

$$\omega(x) = \pi^{1/2}\sigma_{\nu_c}[\nu(x)-\nu_c]. \tag{57}$$

On changing the variable in (47) from ν to ω by means of (52), we have

$$K(x) = \frac{2}{\pi^{1/2}} \cdot \frac{1}{\kappa\sigma_{\nu_c}}\int_{\omega(x)}^{\infty} \sigma_\nu^2\, d\omega. \tag{58}$$

If $0 \leqslant x \leqslant (1+\eta_{\nu_c})^{-1}$, then $\nu(x) = \nu_c$ and $\omega(x) = 0$. Hence by (53)

$$K(x) \simeq \frac{2}{\pi^{1/2}} \eta_{\nu_c} \left\{ \int_0^{\omega_0} e^{-2\omega^2} d\omega + \frac{a^2}{\pi} \int_{\omega_0}^{\infty} \frac{d\omega}{\omega^4} \right\}$$

$$= \eta_{\nu_c} \{ 2^{-1/2} P[\sqrt{2}\omega_0] + \tfrac{2}{3}\pi^{-3/2} a^2 \omega_0^{-3} \}, \tag{59}$$

where $P(x)$ is the probability integral defined by

$$P(x) = \frac{2}{\pi^{1/2}} \int_0^x e^{-t^2} dt. \tag{60}$$

For the value of a given by (55), $P[\sqrt{2}\omega_0]$ differs from unity by less than 10^{-5} and $\frac{2}{3}\pi^{-3/2} a^2 \omega_0^{-3} = 0{\cdot}63 \times 10^{-6}$. Thus $K(x)$ is approximately $2^{-1/2}\eta_{\nu_c}$.

If $(1+\eta_{\nu_c})^{-1} < x \leqslant x_0$, where x_0 is defined by

$$\omega_0 = \pi^{1/2} \sigma_{\nu_c} [\nu(x_0) - \nu_c], \tag{61}$$

then $0 < \omega(x) \leqslant \omega_0$ and, by (58), (59) and (53),

$$K(x) \simeq \eta_{\nu_c} \{ 2^{-1/2} P[\sqrt{2}\omega_0] + \tfrac{2}{3}\pi^{-3/2} a^2 \omega_0^{-3} \} - \frac{2}{\pi^{1/2}} \eta_{\nu_c} \int_0^{\omega(x)} e^{-2\omega^2} d\omega$$

$$= \eta_{\nu_c} \{ 2^{-1/2} P[\sqrt{2}\omega_0] - 2^{-1/2} P[\sqrt{2}\omega(x)] + \tfrac{2}{3}\pi^{-3/2} a^2 \omega_0^{-3} \}. \tag{62}$$

By (49), (61) and (53),

$$x_0 = (1 + \eta_{\nu(x_0)})^{-1} = (1 + \eta_{\nu_c} e^{-\omega_0^2})^{-1}. \tag{63}$$

For the values given by (55),

$$x_0 = 0{\cdot}001\,3.$$

Thus $K(x)$ decreases rapidly in the very small interval $[(1+\eta_{\nu_c})^{-1}, x_0]$ from about $0{\cdot}7 \times 10^6$ to about $0{\cdot}63$.

If $x_0 < x \leqslant 1$, then $\omega(x) > \omega_0$ and

$$K(x) \simeq \frac{2}{\pi^{1/2}} \eta_{\nu_c} \frac{a^2}{\pi} \int_{\omega(x)}^{\infty} \frac{d\omega}{\omega^4} = \tfrac{2}{3}\pi^{-3/2} \eta_{\nu_c} a^2 [\omega(x)]^{-3}. \tag{64}$$

Now from (53), when $x > x_0$,

$$[\omega(x)]^{-2} = \pi^{1/2} (a\eta_{\nu_c})^{-1} \eta_{\nu(x)},$$

and hence by (49), since $x_0 > (1+\eta_{\nu_c})^{-1}$,

$$[\omega(x)]^{-2} = \pi^{1/2} (a\eta_{\nu_c})^{-1} x^{-1} (1-x). \tag{65}$$

On substituting for $\omega(x)$ in (64) we have

$$K(x) \simeq \tfrac{2}{3}\pi^{-3/4} (a/\eta_{\nu_c})^{1/2} x^{-3/2} (1-x)^{3/2} \qquad (x_0 < x \leqslant 1). \tag{66}$$

For the values (55), $K(x)$ decreases from about $0{\cdot}63$ to 0 in the interval $(x_0, 1)$.

In exactly the same way we get the following approximate values for $K_1(x)$:

$$K_1(x) \simeq P(\omega_0) + 2\pi^{-1} a \omega_0^{-1} \qquad (0 \leqslant x \leqslant (1+\eta_{\nu_c})^{-1}), \tag{67}$$

$$K_1(x) \simeq P(\omega_0) - P[\omega(x)] + 2\pi^{-1} a \omega_0^{-1} \qquad ((1+\eta_{\nu_c})^{-1} < x \leqslant x_0), \tag{68}$$

$$K_1(x) \simeq 2\pi^{-3/4} (a/\eta_{\nu_c})^{1/2} x^{-1/2} (1-x)^{1/2} \qquad (x_0 < x \leqslant 1). \tag{69}$$

For the values (55), $2\pi^{-1} a \omega_0^{-1} = 0{\cdot}002\,4$, $P(w_0) = 0{\cdot}999\,8$. Thus $K_1(x)$ decreases from about 1 to $0{\cdot}002\,4$ in the interval $((1+\eta_{\nu_c})^{-1}, x_0)$ and from $0{\cdot}002\,4$ to 0 in the interval $(x_0, 1)$.

We next need $\int_0^1 K_1(x)\,dx$ and for this it is easiest to use the formula (43). When $\kappa_\nu = \kappa$, this gives

$$\int_0^1 K_1(x)\,dx = \int_0^\infty \frac{\sigma_\nu}{1+\eta_\nu} d\nu \simeq \int_{-\infty}^{\infty} \frac{\sigma_\nu}{1+\eta_\nu} d\nu.$$

By (52) and (53)

$$\int_0^1 K_1(x)\,dx \doteq \frac{2}{\pi^{1/2}}\left\{\int_0^{\omega_0}\frac{d\omega}{e^{\omega^2}+\eta_{\nu_c}}+\int_{\omega_0}^{\infty}\frac{d\omega}{\pi^{1/2}a^{-1}\omega^2+\eta_{\nu_c}}\right\}$$
$$=\frac{2}{\pi^{1/2}}\int_0^{\omega_0}\left\{\frac{1}{e^{\omega^2}+\eta_{\nu_c}}-\frac{1}{\pi^{1/2}a^{-1}\omega^2+\eta_{\nu_c}}\right\}d\omega+\pi^{1/4}\left(\frac{a}{\eta_{\nu_c}}\right)^{1/2}.$$

The integral is of the order of $\eta_{\nu_c}^{-2}$ and is therefore negligible compared with the second term. Hence

$$\int_0^1 K_1(x)\,dx \doteq \pi^{1/4}(a/\eta_{\nu_c})^{1/2}.$$

Following Sobolev, we shall write

$$A=\left(\int_0^1 K_1(x)\,dx\right)^{1/2}\doteq \pi^{1/8}(a/\eta_{\nu_c})^{1/4}. \tag{70}$$

For the values (55),

$$A\doteq 0{\cdot}01.$$

7. *The function* $H(x)$.—This is the solution of (30), or, by (**2**), p. 107, Corollary 2, of

$$\frac{1}{H(x)}=A+\frac{1}{2}\int_0^1\frac{x'H(x')K(x')}{x+x'}\,dx'. \tag{71}$$

As for all the H-functions, $H(0)=1$ and $H(x)$ is continuous and steadily increasing in the interval (0, 1). From (71) it follows that

$$1/H(x)>A$$

and therefore that

$$1\leqslant H(x)<A^{-1} \quad \text{for} \quad 0\leqslant x\leqslant 1. \tag{72}$$

Sobolev gives the following approximate formula for $H(x)$:

$$H(x)\doteq 1+\frac{2x/A^2}{1+2x/A}. \tag{73}$$

I have failed to find adequate justification for this approximation, which is stated without proof. It makes $H(0)=1$ and $H(x)$ a continuous and steadily increasing function in (0, 1), but it makes $H(1)>A^{-1}$, which contradicts (72). Moreover the derivative of $H(x)$ is infinite at $x=0$, whereas that given by (73) is $2/A^2$. If (73) does give a sufficiently good approximation for $H(x)$, then

$$H(A)>\tfrac{2}{3}A^{-1}. \tag{74}$$

In discussing the magnitudes of the terms in (46), I shall only use (74) to show that, if Sobolev's approximation (73) is satisfactory, then the term which he neglects is not negligible. If we work solely with the wider inequality (72), we can only show that the term may not be negligible.

8. *A discussion of the solution* (46).—Sobolev's solution is (46) with $\kappa_\nu=\kappa$, $b_\nu^{(0)}=B$, $b_\nu^{(1)}=0$, and in his deductions he neglects the term

$$\int_0^1\frac{x'H(x')K_1(x')}{x+x'}\,dx'. \tag{75}$$

I shall show that if (73) is a satisfactory approximation, then (75) is of the same order of magnitude as $A=\left(\int_0^1 K_1(x)\,dx\right)^{1/2}$, the term retained by Sobolev. If we can only use the inequality (72), the proof that (75) is not larger than A still holds, but the proof that it is not much smaller breaks down.

Since all the functions appearing in (75) are positive, the integral is greatest when $x=0$ and we therefore want the size of

$$\int_0^1 H(x')K_1(x')\,dx'.$$

By (72) and (70), $$\int_0^1 H(x')K_1(x')\,dx' < A^{-1}\int_0^1 K_1(x')\,dx' = A. \tag{76}$$

Hence (75) is not greater than A.

When a and η_{ν_c} have the values given by (55), then $x_0 = 0{\cdot}001\,3$ and $A = 0{\cdot}01$, and so

$$A > x_0. \tag{77}$$

From (63) and (70) it is easily seen that (77) will be true for any fixed a (and therefore fixed ω_0) provided η_{ν_c} is large enough. We shall assume that this is the case.

We have

$$\int_0^1 H(x')K_1(x')\,dx' \geqslant \int_A^1 H(x')K_1(x')\,dx' \geqslant H(A)\int_A^1 K_1(x')\,dx',$$

since $H(x')$ is steadily increasing in $(A, 1)$. By (77) and (69)

$$\begin{aligned}\int_0^1 H(x')K_1(x')\,dx' &\geqslant H(A)\,.\,2\pi^{-3/4}(a/\eta_{\nu_c})^{1/2}\int_A^1 x^{-1/2}(1-x)^{1/2}\,dx\\ &= A^2H(A)\,.\,\frac{2}{\pi}\left\{\frac{\pi}{2} - \sin^{-1}A^{1/2} - A^{1/2}(1-A)^{1/2}\right\}\\ &\geqslant \tfrac{1}{2}A^2H(A),\end{aligned}$$

since $A \leqslant 0{\cdot}1$ in all the cases we are considering. If (74) is true, then it follows that

$$\int_0^1 H(x')K_1(x')\,dx' \geqslant \tfrac{1}{3}A.$$

Thus the term is not negligible if Sobolev's approximation holds. Since $H(A) \geqslant 1$ (by (72)), the term lies in any case between A and $\frac{1}{2}A^2$.

We shall now investigate the other integrals appearing in (46) in a similar way. We shall, however, only obtain an upper bound for each integral.

The integral

$$\int_0^1 \frac{x'^2H(x')K_1(x')}{x+x'}\,dx'$$

is less than (75) and is therefore less than A.

The integral

$$\int_0^1 x'H(x')K(x')\,dx' \tag{78}$$

requires more detailed consideration. By (72),

$$\int_0^1 x'H(x')K(x')\,dx' \leqslant A^{-1}\int_0^1 x'K(x')\,dx'. \tag{79}$$

On substituting from (31) and inverting the order of integration, we have

$$\begin{aligned}\int_0^1 x'K(x')\,dx' &= \kappa^{-1}\int_0^\infty \sigma_{\nu'}{}^2\,d\nu' \int_0^{(1+\eta_{\nu'})^{-1}} x'\,dx'\\ &= \tfrac{1}{2}\kappa^{-1}\int_0^\infty \frac{\sigma_{\nu'}{}^2}{(1+\eta_{\nu'})^2}\,d\nu'\\ &\leqslant \tfrac{1}{2}\kappa^{-1}\int_{-\infty}^\infty \frac{\sigma_{\nu'}{}^2}{(1+\eta_{\nu'})^2}\,d\nu'.\end{aligned}$$

By (52) and (53)

$$\int_0^1 x'K(x')\,dx' \simeq \pi^{-1/2}\eta_{\nu_c}\left\{\int_0^{\omega_0}\frac{d\omega}{(e^{\omega^2}+\eta_{\nu_c})^2}+\int_{\omega_0}^{\infty}\frac{d\omega}{(\pi^{1/2}a^{-1}\omega^2+\eta_{\nu_c})^2}\right\}$$

$$=\pi^{-1/2}\eta_{\nu_c}\int_0^{\omega_0}\left\{\frac{1}{(e^{\omega^2}+\eta_{\nu_c})^2}-\frac{1}{(\pi^{1/2}a^{-1}\omega^2+\eta_{\nu_c})^2}\right\}d\omega$$

$$+\tfrac{1}{4}\pi^{1/4}(a/\eta_{\nu_c})^{1/2}.$$

The first term is of the order of $\eta_{\nu_c}^{-2}$ and is negligible compared with the second term, which, by (70), is $\frac{1}{4}A^2$. Hence

$$\int_0^1 x'K(x')\,dx' \simeq \tfrac{1}{4}A^2$$

and (79) gives

$$\int_0^1 x'H(x')K(x')\,dx' \leqslant \tfrac{1}{4}A. \tag{80}$$

From these results we see that no integral in (46) is greater than A, but, on the other hand, there is no reason to suppose that any one is negligible compared with A. In view of this we shall now consider whether Sobolev's assertion about the central intensity of an absorption line is justified.

When $\nu=\nu_c$ then $x=\mu(1+\eta_{\nu_c})^{-1}$ and this is very small, but it does not necessarily follow that $H(x)\simeq H(0)=1$, because $H(x)$ increases very rapidly near $x=0$. For this reason we shall only consider the case in which $x=0$.

Since $\lambda_{\nu_c}=(1+\eta_{\nu_c})^{-1}$, this is negligible compared with A, and (46) gives

$$I_{\nu_c}(0,0) \simeq b_{\nu_c}^{(0)}\left\{A-\tfrac{1}{2}\int_0^1 H(x')K_1(x')\,dx'\right\}+\tfrac{1}{2}b_{\nu_c}^{(1)}\left\{\int_0^1 x'H(x')K(x')\,dx'\right.$$

$$\left.+\int_0^1 x'H(x')K_1(x')\,dx'\right\}. \tag{81}$$

To the same order of approximation (16) gives (with an obvious notation)

$$I_{\nu_c}^{(\mathrm{coh})}(0,0) \simeq b_{\nu_c}^{(0)}\lambda_{\nu_c}^{1/2} \simeq b_{\nu_c}^{(0)}\eta_{\nu_c}^{-1/2}.$$

Now by (76) and (70)

$$A-\tfrac{1}{2}\int_0^1 H(x')K_1(x')\,dx' > \tfrac{1}{2}A \simeq \tfrac{1}{2}\pi^{1/8}(a/\eta_{\nu_c})^{1/4}.$$

Hence

$$I_{\nu_c}(0,0)/I_{\nu_c}^{(\mathrm{coh})}(0,0) > \tfrac{1}{2}\pi^{1/8}(a\eta_{\nu_c})^{1/4} \simeq 5$$

for the values given by (55). Thus the central intensity of a line at the limb is much lighter on the hypothesis of non-coherent scattering than on that of coherent scattering. A similar result can be expected to hold when $\mu>0$, but it is difficult to establish this without computing $H(x)$.

In conclusion I should like to thank Professor H. H. Plaskett, Dr M. G. Adam and Mr D. W. N. Stibbs for their interest and for helpful criticisms.

APPENDIX

The scattering function found by Chandrasekhar's method

We consider the same scattering atmosphere as in Section 4 subject to the same incident radiation. In this section it is convenient to denote inwardly directed intensities by the superscript $-$ and outwardly directed intensities by the superscript $+$. Then the intensity of the incident radiation is, by (20),

$$I_\nu^-(0, -\mu') = \tfrac{1}{2}F_{\nu_0}\delta(\nu_0-\nu')\delta(\mu_0-\mu'); \tag{82}$$

the law of diffuse reflection is given by (21), and the intensity of the diffusely reflected radiation is

$$I_\nu^+(0, \mu) = \tfrac{1}{4}F_{\nu_0}S(\nu, \nu_0, \mu, \mu_0)/\mu. \tag{83}$$

Because t_ν varies with ν, it is necessary to work with the depth z and we shall therefore write $I_\nu(z, \mu)$ where we have had $I_\nu(t_\nu, \mu)$ hitherto. The equation of transfer for $I_\nu(z, \mu)$ is (17) without the term $\kappa_\nu B_\nu$, viz.

$$\mu\frac{dI_\nu(z, \mu)}{\rho\, dz} = (\kappa_\nu+\sigma_\nu)I_\nu(z, \mu) - \tfrac{1}{2}\sigma_\nu\int_0^\infty \sigma_{\nu'}d\nu'\int_{-1}^{1} I_{\nu'}(z, \mu')\, d\mu'. \tag{84}$$

In this we have used (18).

We now define the intensity of the reduced incident radiation by

$$I_\nu^R(z, \mu) = \tfrac{1}{2}F_{\nu_0}\delta(\nu_0-\nu)\delta(\mu_0+\mu)e^{-\rho(\kappa_{\nu_0}+\sigma_{\nu_0})z/\mu_0}. \tag{85}$$

Then $I_\nu^R(z, \mu)\equiv 0$ except when $\nu=\nu_0$, $\mu=-\mu_0$, and it satisfies the transfer equation

$$\mu\frac{dI_\nu^R(z, \mu)}{\rho\, dz} = (\kappa_\nu+\sigma_\nu)I_\nu^R(z, \mu). \tag{86}$$

Let

$$\mathfrak{I}_\nu(z, \mu) = I_\nu(z, \mu) - I_\nu^R(z, \mu). \tag{87}$$

Then $\mathfrak{I}_\nu(z, \mu)$ is the intensity in the reduced radiation field, and by (84) and (86) it satisfies the equation

$$\mu\frac{d\mathfrak{I}_\nu(z, \mu)}{\rho\, dz} = (\kappa_\nu+\sigma_\nu)\mathfrak{I}_\nu(z, \mu) - \sigma_\nu\mathfrak{J}(z), \tag{88}$$

where

$$\mathfrak{J}(z) = \frac{1}{2}\int_0^\infty \sigma_{\nu'}\, d\nu'\int_{-1}^{1} I_{\nu'}(z, \mu')\, d\mu'. \tag{89}$$

Also, by (82), (85) and (87), the boundary condition for $\mathfrak{I}_\nu(z, \mu)$ is

$$\mathfrak{I}_\nu^-(0, -\mu')\equiv 0, \tag{90}$$

and by (83), (85) and (87)

$$\mathfrak{I}_\nu^+(0, \mu) = I_\nu^+(0, \mu) = \tfrac{1}{4}F_{\nu_0}S(\nu, \nu_0, \mu, \mu_0)/\mu. \tag{91}$$

From (89, (82), (83), we have

$$\begin{aligned}\mathfrak{J}(0) &= \tfrac{1}{4}F_{\nu_0}\int_0^\infty \sigma_{\nu'}\delta(\nu_0-\nu')\, d\nu'\int_0^1 \delta(\mu_0-\mu')\, d\mu' \\ &\quad + \tfrac{1}{8}F_{\nu_0}\int_0^\infty \sigma_{\nu'}\, d\nu'\int_0^1 S(\nu', \nu_0, \mu', \mu_0)\frac{d\mu'}{\mu'} \\ &= \tfrac{1}{4}F_{\nu_0}\sigma_{\nu_0}\left\{1 + \frac{1}{2}\int_0^\infty \sigma_{\nu'}{}^2 d\nu'\int_0^1 \frac{S(\nu', \nu_0, \mu', \mu_0)}{\sigma_{\nu'}\sigma_{\nu_0}}\frac{d\mu'}{\mu'}\right\}\end{aligned} \tag{92}$$

by the properties of the δ-function.

If we put $z=0$ in (88) and distinguish between upward and downward intensities, we have, by (90) and (91),

$$\left[\frac{d\mathfrak{F}_\nu^+(z,\ \mu)}{dz}\right]_{z=0} = \frac{\rho}{\mu}\left\{\frac{1}{4\mu}F_{\nu_0}(\kappa_\nu+\sigma_\nu)S(\nu,\ \nu_0,\ \mu,\ \mu_0)-\sigma_\nu\mathfrak{J}(0)\right\}, \tag{93}$$

$$\left[\frac{d\mathfrak{F}_{\nu'}^-(z,-\mu')}{dz}\right]_{z=0} = \frac{\rho}{\mu'}\sigma_{\nu'}\mathfrak{J}(0). \tag{94}$$

Now by the principle of invariance, the law of diffuse reflection is unchanged if we omit a layer of the atmosphere of thickness z, and since no component of the field $I_\nu(z,\mu)$ is being generated within the atmosphere itself, it follows from (21) that

$$I_\nu^+(z,\mu)=\int_0^\infty d\nu'\int_0^1 I_{\nu'}^-(z,\ -\mu')\frac{S(\nu,\nu',\mu,\mu')}{2\mu}\,d\mu'. \tag{95}$$

By (87), (85) and the properties of the δ-function, this gives

$$\mathfrak{F}_\nu^+(z,\mu)=\int_0^\infty d\nu'\int_0^1 \mathfrak{F}_{\nu'}^-(z,\ -\mu')\frac{S(\nu,\nu',\mu,\mu')}{2\mu}\,d\mu' + \frac{1}{4\mu}F_{\nu_0}S(\nu,\ \nu_0,\ \mu,\ \mu_0)e^{-\rho(\kappa_{\nu_0}+\sigma_{\nu_0})z/\mu_0}. \tag{96}$$

On differentiating this with respect to z and putting $z=0$, we have, by (93) and (94),

$$\frac{\rho}{\mu}\left\{\frac{1}{4\mu}F_{\nu_0}(\kappa_\nu+\sigma_\nu)S(\nu,\ \nu_0,\ \mu,\ \mu_0)-\sigma_\nu\mathfrak{J}(0)\right\} = \frac{\rho\mathfrak{J}(0)}{2\mu}\int_0^\infty\sigma_{\nu'}d\nu'\int_0^1 S(\nu,\ \nu',\ \mu,\ \mu')\frac{d\mu'}{\mu'}-\frac{\rho}{4\mu\mu_0}(\kappa_{\nu_0}+\sigma_{\nu_0})F_{\nu_0}S(\nu,\ \nu_0,\ \mu,\ \mu_0),$$

and this can be written

$$\tfrac{1}{4}F_{\nu_0}S(\nu,\ \nu_0,\ \mu,\ \mu_0)\left\{\frac{\kappa_\nu+\sigma_\nu}{\mu}+\frac{\kappa_{\nu_0}+\sigma_{\nu_0}}{\mu_0}\right\} = \sigma_\nu\mathfrak{J}(0)\left\{1+\frac{1}{2}\int_0^\infty\sigma_{\nu'}^2\,d\nu'\int_0^1\frac{S(\nu,\nu',\mu,\mu')}{\sigma_\nu\sigma_{\nu'}}\frac{d\mu'}{\mu'}\right\}. \tag{97}$$

Finally, on substituting for $\mathfrak{J}(0)$ from (92), we get (22), after which the analysis of Section 4 follows unchanged.

St. Hugh's College,
Oxford:
1952 *September* 30.

References

(1) V. A. Ambartsumian, *C. R. (Doklady) Acad. Sci. U.R.S.S.*, **38**, 257, 1943.
(2) S. Chandrasekhar, *Radiative Transfer*, Oxford, 1950.
(3) V. Kourganoff, *Basic Methods in Transfer Problems*, Oxford, 1952.
(4) A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms*, Cambridge, 1934.
(5) G. Münch, *Ap. J.*, **109**, 275, 1949.
(6) M. P. Savedoff, *Ap. J.*, **115**, 509, 1952.
(7) V. V. Sobolev, *R.A.J.*, **26**, 129, 1949.
(8) P. ten Bruggencate and J. Houtgast, *Z. f. Ap.*, **20**, 149, 1940.

QUARTZ CLOCKS OF THE GREENWICH TIME SERVICE

Humphry M. Smith

(Communicated by the Astronomer Royal)

(Received 1952 November 13)

Summary

A general account is given of the development of the quartz clocks of the Greenwich Time Service, and the present installation is described. The methods by which clock performance may be studied are described and a convenient criterion is proposed. Quantitative results are presented in tabular form. The degree of stability now attained represents a significant advance upon that of some five years ago. Details are given of development work now in progress with a view to improving stability still further.

---

1. *General description.*—The first quartz clock of the Observatory was installed at Greenwich in 1939, and was of the pattern developed at the National Physical Laboratory (**1**). This employed a ring crystal supported on phosphor-bronze pins and mounted in an evacuated glass vessel held at a constant temperature by means of a mercury thermometer. The crystal was incorporated in a simple Pierce oscillator circuit, and was designed to oscillate at 99·3 kc/s. The divider chain performed successive division by 3, 6 and 11, giving an output of 500 cycles per sidereal second to drive a phonic motor. This clock, Q3, took its place among the standard pendulum clocks of the Time Service, which were also supplemented by data received daily from the National Physical Laboratory regarding two similar quartz clocks Q2 and Q6.

Clocks employing GT-cut plate crystals in a drive circuit based upon the Meacham bridge (**2**) were developed independently at the Post Office Radio Branch Laboratories. The oven and temperature-control arrangements were essentially a simplified version of the thermostat designed by Turner (**3**). The fundamental crystal frequency was 100 kc/s, which was converted by means of regenerative dividers (**4**) into a 1 kc/s output used to drive a phonic motor. A group of three such oscillators, known as "Group 4" (**5**) attained a high standard of performance, and in 1942 arrangements were made for information relating to these clocks to be communicated daily to the Royal Observatory, where it proved of the greatest value. Within a year it became clear that the long-term performance of these clocks was so markedly superior to even the best of the pendulum clocks available to the Time Service that the pendulum clock was rendered virtually obsolete as a long-term standard (**6**).

It was clearly desirable that similar quartz clocks should be installed at the Observatory, so that the Service would not be entirely dependent upon primary standards at an external establishment. There was also an urgent need for secondary standards having good day-to-day stability for the control of the signals, and in this respect also the superiority of the quartz clock had been established. The first three clocks supplied by the Post Office were of the Group 4 pattern, and were brought into use early in 1944 for the control of the GBR radio time signals, thus bringing about a considerable improvement in

the day-to-day uniformity of these time signals. Unfortunately, the value of these clocks as long-term standards was seriously impaired by the fact that they were mains operated and that, under war-time conditions, interruptions of the mains electricity supply were of frequent occurrence. It was thus still necessary to base the service on the Post Office Group 4 standards.

During 1945 four groups of three oscillators, similar in general arrangements to the Post Office Group 4, were installed in new clock cellars at the Observatory station at Abinger (**7**). These oscillators, together with their ancillary equipment, were designed for battery operation, and an extensive installation of batteries and control equipment, together with a stand-by diesel generator, was provided. This overcame the main cause of disturbance to the oscillators, and longer uninterrupted clock runs were made possible. The temperature-control circuits, however, were still dependent on a 50 c/s a.c. supply, and in the event of mains failures this was provided automatically by a battery-driven motor alternator. Two similar groups of three oscillators each were subsequently installed at Greenwich, and incorporated some minor improvements. Special types of phonic motors were developed in order that all requirements of the service could be provided directly from the quartz clocks (**8**). Motors incorporating sidereal gearing were used to provide sidereal seconds for the astronomical time observations; and more elaborate timing devices, capable of producing the various forms of time signal distributed by the Department, made the full accuracy of the quartz clocks available for direct practical use. It had thus become possible for the quartz clocks to perform, with greater accuracy, all the functions hitherto carried out by the pendulum clocks: in fact, the manifest inferiority of the pendulum clocks led to their complete elimination.

One result of the widespread employment of quartz clocks was that the clock comparison equipment which had been adequate for use with pendulum clocks was found to be incapable of meeting the new demands for accuracy. In particular, it became necessary to replace the electro-mechanical chronographs by fully electronic decimal counter chronometers (**9**). Moreover, quartz clocks may be conveniently intercompared with great accuracy by methods which are inapplicable to pendulum clocks. Relative rates may be readily obtained by means of a precision beat comparator (**10**) and integrated time differences between standards by means of rotary beat counters (**11**). It thus becomes possible to maintain a constant check on the performance of all oscillators and to connect the dividers and phonic motors to those oscillators giving the best performance at the time.

As initially installed, the three oscillators of each group were mounted on one rack, in accordance with standard Post Office practice. Although oscillators employing GT-cut crystals are not unduly sensitive to mechanical disturbance, it was decided to re-mount the oscillators singly on individual brick and concrete piers, thus making it possible to carry out work on one oscillator without risk of disturbance to the others of the group. In the course of the necessary re-wiring, the power supplies were re-arranged to permit convenient isolation of each oscillator. New crystal ovens were constructed in the Observatory workshop, to an improved Post Office specification which is designed to accommodate either GT or ring crystals. The oven is located in a hollow in a concrete block on top of the pier, where it is protected from accidental disturbance and draughts. The temperature-control unit and maintaining amplifier are mounted independently above the oven. Some minor changes were made in the circuit

arrangements: high-stability components were employed, and the layout was modified to facilitate servicing. The step-up ratio of the transformer in the temperature-control circuit was increased to improve the sensitivity, and the constant-temperature arms of the control bridge were re-wound on a solid brass rod. The period of the oven-control cycle was thereby reduced to about five seconds. Duplicate thyratrons were fitted, with separate signal lamps, thus giving warning of impending thyratron failure, and permitting replacement without disturbance of the temperature control. In the new maintaining amplifiers toroidal transformers are employed and precautions have been taken to render the amplifier gain still less sensitive to variations in supply voltages or in the loading of the 100 kc/s output. Some amplifiers have been fitted with prototype long-life valves. During the testing and setting-up of these oscillators it became necessary to devise new techniques for alignment and adjustment.

In view of the excellent results achieved with a bridge drive circuit in association with GT-cut plate crystals, the Post Office Radio Branch Laboratories proceeded to produce a bridge circuit suitable for a ring crystal. Three oscillators, forming the Post Office Group 5, employed ring crystals supported by nickel pins. Later oscillators, forming Group 9, utilized softer metal for the crystal supports, and in the latest type the ring is suspended on silk threads. Very high values of Q (i.e. sharpness of resonance) have been obtained, but perhaps the most significant characteristic of the new ring oscillators is the substantial reduction of the initial frequency drift. An oscillator of this type was installed at the Observatory in 1950 May, and was brought into use almost immediately as the working standard E5. For subsequent employment as a long-term standard, its useful run began only a week or so after installation. A remarkable feature of this particular standard is the small value of the frequency drift, which amounts to only 1 part in 10^8 per year. A second oscillator of the same type, E6, was installed a few months later, and has exhibited a similar degree of stability, but is characterized by a uniformly changing frequency drift. Two further ring-crystal oscillators were installed towards the end of 1951. One of these, C5, is performing satisfactorily, but the other, C6, failed to reach the high standard of the other ring oscillators. It was accordingly removed from service, and the maintaining amplifier was modified in the light of more recent experience. The new run of this oscillator shows a satisfactory improvement in short-term stability, and there is good reason to suppose that the long-term performance will be correspondingly improved.

The ancillary equipment associated with the oscillators has also received attention. The regenerative dividers providing an output at 1 kc/s have been replaced by improved models developed at the Royal Observatory. The earlier pattern employed a large number of metal rectifiers and a total of fourteen transformers. In the new pattern, no metal rectifiers are employed, electronic mixing is used, and the number of transformers has been reduced to four. The valves were previously operated at nearly full rated anode current, and performance was adversely affected by valve ageing. The valves are now operated at low anode current, with consequent saving of power and reduced sensitivity to changes in valve characteristics. In order to meet the need for accurately spaced seconds pulses, new pulse dividers have been designed. A series of identical plug-in units, each giving one output pulse for every ten input pulses, may be connected in cascade to perform frequency division by any required power of ten. The units function with equal reliability at any input frequency

from 100 000 pulses per second to one pulse per second. A series consisting of five such units supplied with pulses derived from the 100 kc/s crystal frequency provides seconds pulses of more than adequate accuracy. Full advantage is thus taken of the extremely high stability of the latest crystal oscillators.

2. *Assessment of long-term performance.*—For the operation of a National Time Service of high accuracy the clocks should provide a time scale that is uniform

(*a*) from month to month over a period of two years or so;

(*b*) from day to day over a period of one month or longer.

The clocks which adequately fulfil the first requirement are employed for long-term prediction. Those with a high degree of day-to-day stability are used as working standards in the control of time-signal transmissions. It is therefore necessary to assess the quality of the available clocks from these two different aspects.

The problem of relating the performance of a modern quartz clock to the ultimate standard of time, the rotation of the Earth, has become increasingly difficult on account of the latest developments in quartz clocks. For some years the uniformity of the time scale provided by the clocks has surpassed that defined by the Earth: this disparity has been growing more pronounced since the introduction of improved clocks employing ring-crystal oscillators (**12**). In an attempt to remove the effects of one known cause of irregularity in observed time, it has been the practice for the last five years at the Royal Greenwich Observatory to apply current corrections for the variation in longitude arising from polar motion, or more briefly, "polar variation". These corrections are based on independent data derived from variations of latitude observed at the U.S. Naval Observatory. Another known cause of irregularity is the annual fluctuation in the rate of rotation of the Earth. The mean amplitude of the fluctuation over a number of years was independently determined by several investigators. The degree of agreement between these various results suggested that the effect was sensibly constant from year to year, and that considerable advantage would be gained by the application of a correction based on the average fluctuation observed in preceding years. It has since become apparent that the annual fluctuation is subject to appreciable variation from year to year, and that the use of an average value may be misleading (**13**). This fluctuation can only be determined from the time observations themselves, and current estimations of its magnitude are subject to considerable uncertainty. It has in fact become impracticable to obtain a sufficiently uniform time scale by applying to the astronomical time observations provisional corrections for polar variation and annual fluctuation. A time scale of adequate uniformity may however be established by taking the mean of a number of specially selected quartz clocks of the highest precision. But such a scale is not absolute unless it can be related to the rotation of the Earth. It is therefore necessary to calibrate the scale with reference to the general trend of the astronomical observations over a period of two or three years. The procedure now adopted at the Royal Greenwich Observatory is to use a mean of the best clocks at the various cooperating establishments to define a time scale that is, as nearly as possible, uniform, and to determine the rate of the mean clock by consideration of the astronomical time observations made over an extended period at Greenwich and Abinger.

The application in practice of these principles may be illustrated by reference to the analysis performed for the period 1950 June to 1952 June. Information

relating to all available clocks was readily accessible from the current graphs of clock intercomparisons. An examination of the performance of the clocks during the period under review permitted the selection of the five best clocks. These clocks were then referred to the astronomically observed time corrected for polar variation and the provisionally adopted annual fluctuation. An adequate basis was thus provided for the adoption, for each clock, of a preliminary parabolic ephemeris extending over the two-year interval. Where the intercomparisons gave clear evidence of a change of rate or drift in any particular clock, due allowance was made. In the course of the two years, two of the clocks, E5 and 9A, required no such adjustment; the remaining three each had one apparent change of drift: 9C in 1951 April, Q13 in 1951 May, and EA in 1951 November. Each clock was then referred to the monthly means of the observations, corrected for polar variation but not for annual fluctuation. Corresponding monthly points in successive years were differenced to give a series of annual rates. If the fluctuation in the rate of rotation of the Earth were strictly repetitive from one year to the next, it would have no effect upon clock rates determined over these

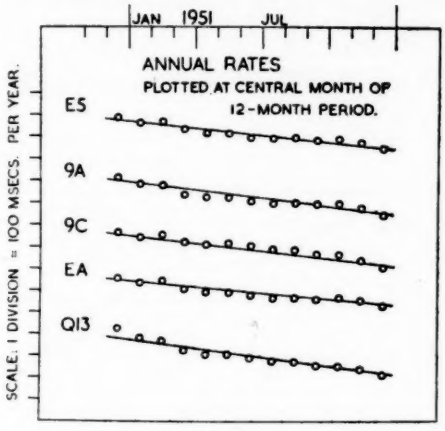

FIG. 1.

twelve-monthly periods. In practice it is found that the fluctuation varies somewhat in amplitude and phase, but that this procedure is effective in removing the major part of the fluctuation in determined rate. Moreover, the use of a long interval considerably reduces the residual effect, and largely eliminates errors arising from observational scatter. The efficacy of the process is demonstrated by the fact that the plotted points adhere closely to straight lines (see Fig. 1), which were used to determine corrections to the parameters of the preliminary parabolic ephemerides.

The five clocks were then combined into a mean clock, which established a provisional uniform time system of which the uniformity and scale value depended upon the drift and rate adopted for the mean clock. Since the drift and rate of each individual clock had been ultimately determined by consideration of annual rates, given by the observations, the time system defined by the mean clock was fundamentally related to the astronomical standard of time, from which the annual fluctuation had been removed. The system was however largely independent of the particular values adopted for the amplitude and phase of

the annual fluctuation. This time system is therefore suitable as a basis for the evaluation of the 12-month and 6-month periodicities of the annual fluctuation by a series of simple harmonic analyses, each covering twelve successive monthly means of observations corrected for polar variation. It has been found convenient to start each harmonic analysis at the beginning of a quarter-year so that any given analysis has an overlap of nine months with the previous and following analysis. There were therefore five harmonic analyses in the two-year period under consideration, the first and fifth being independent of each other, and the others overlapping. The five results for the annual fluctuation appeared to be in reasonable agreement, and a convenient set of values was adopted for the whole period. Corrections for annual fluctuation computed from these were then applied to the monthly means of observations, bringing the plotted points to the neighbourhood of a horizontal straight line. The mean ordinate was computed and applied, fixing the zero-point of the uniform time system. The uniformity and scale value having been fixed as previously described, the time system was thus completely determined. The remaining scatter was attributed to observational error, and amounted to a mean deviation of $\pm 6{\cdot}2$ milliseconds for a single monthly point.

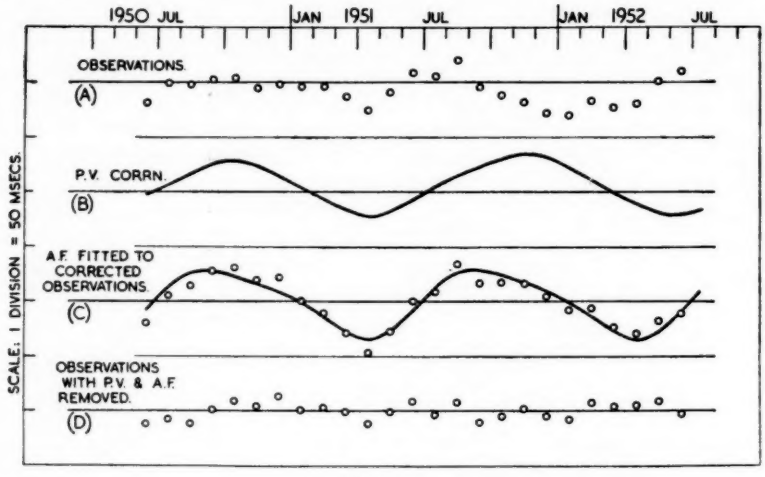

FIG. 2.

In Fig. 2 are shown

(A) the uncorrected observations (monthly means) referred to the mean clock;

(B) the correction for polar variation based on latitude observations made in the U.S.A.;

(C) the observations corrected for polar variation, but still subject to annual fluctuation, together with the curve based upon the mean annual fluctuation adopted for the period; and

(D) the observations corrected for polar variation and annual fluctuation, with the mean ordinate removed.

It will be noticed that the polar variation correction, Fig. 2 (B), shows a preponderance of positive values, arising from an apparent discrepancy in the

adopted mean latitude of the U.S. observing station. This is not of serious importance, as it affects only the zero-point of the uniform time system, and thus has no effect whatever upon the determined annual fluctuation, the determined astronomical time (G.M.T.), or the assessment of clock performance. An adjustment of the adopted mean latitude would give rise to a discontinuity in the uniform time system, and is thus better delayed until a revised mean can be adopted from, say, ten years of results. For the same reason, changes in the adopted position of the Earth's mean pole, which are equivalent to adjustments of adopted mean latitude for many stations, are best made as infrequently as possible. It is of interest to note that during the period considered the effect of the polar variation on time observations at Greenwich was nearly equal in amplitude and opposite in phase to the effect of the annual fluctuation. Furthermore, any errors in the adopted corrections for polar variation will be reflected in the determined annual fluctuation. It has been found that the corrections derived from the U.S. latitude observations are substantially confirmed by those based upon the results published subsequently by the International Latitude Bureau.

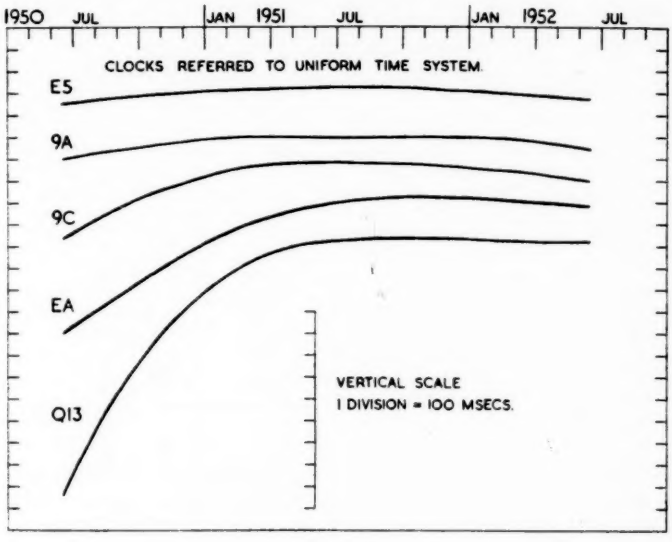

FIG. 3.

The establishment of a uniform time system has been described in some detail, since this is of fundamental importance in the assessment of long-term clock performance (Fig. 3). The mid-monthly values of the daily rates of typical clocks, in terms of this uniform time system, are listed in Table I. The stability in rate from month to month of the two best clocks, E5 and 9A, is indicated in Fig. 4. The clock 9A has been in continuous operation since 1948 January, and has evidently settled to a practically uniform frequency drift. The clock E5, on the other hand, was not installed until 1950 May, and appears to have settled immediately to a frequency drift which was not only practically uniform but also very small.

The most important requirement in a long-period standard clock is that it should run uniformly with a constant frequency drift for as long a period as possible. Twelve months of uniform run are needed before even one determination of annual rate can be made. A uniform run of two years provides twelve monthly values of annual rate, and represents the shortest period that can give a reliable indication of frequency drift. For any period of approximately uniform run, the first differences of the monthly values of rate will represent the monthly drift in the rate of the clock. For a perfect clock the first differences will be constant, and the second differences zero. The magnitude of the second difference will indicate the departure of the clock from uniformity of run. It was therefore decided to evaluate the mean absolute monthly second difference for each clock as a numerical criterion of clock performance. The appropriate criterion values, in milliseconds per day per month per month, are quoted at the foot of Table I. The requirements of a modern time service are such that a clock should maintain a criterion value not exceeding 0·10 for two years or so.

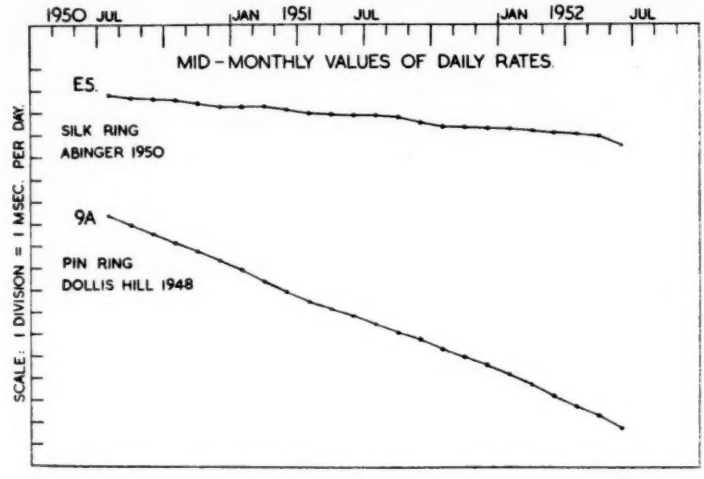

FIG. 4.

It may be noted that the criterion so determined is largely independent of errors in the adopted rate and drift of the reference time system. In fact, over an isolated period of two years, it may be regarded as independent of the astronomical determinations of time, since, for such periods, the mean of several modern quartz clocks gives a more uniform time scale than that defined by the rotation of the Earth. The astronomical observations relating the clocks to the rotation of the Earth are however essential both in current work and in retrospective assessment of performance. Even the best modern clocks are liable to slight changes of rate and drift which cannot be detected by clock intercomparisons until some time has elapsed and an appreciable error has accumulated. The astronomical observations, although subject to considerable scatter, serve to protect the currently adopted time from gross errors which might otherwise arise on this account. In retrospective assessment, the practice of judging the stability of quartz clocks in terms of their relative performance cannot be extended over periods greatly in excess of two years. In order to

control long-period changes in rate and drift, it is necessary to refer to an independent fundamental standard. The only practicable fundamental standard at present available for this purpose is the rate of rotation of the Earth, as indicated by observations of stellar transits. A more uniform fundamental standard of time would be provided by the period of revolution of the Earth around the Sun.

TABLE I

Mid-monthly values of daily rates of selected clocks (milliseconds per day)

	Abinger		Greenwich		Dollis Hill			N.P.L.
Standard	E5	E6	F1	9A	9C	EA	EB	Q13
Ring or Plate	R	R	P	R	R	R	R	R
Silk, Pin or Clamp Support	S	S	C	P	P	S	S	S
1950								
July	+47·83		+84·54	−26·61	−0·32	−56·13	−41·48	+30·47
August	+47·69		+83·83	−27·03	−0·76	−56·54	−41·56	+29·33
September	+47·61		+83·13	−27·47	−1·22	−56·81	−42·01	+28·17
October	+47·58		+82·37	−27·85	−1·83	−57·24	−42·69	+27·12
November	+47·44	−40·39	+81·53	−28·26	−2·35	−57·81	−43·37	+26·22
December	+47·26	−41·98	+80·71	−28·67	−2·69	−58·36	−43·78	+25·30
1951								
January	+47·24	−43·32	+80·04	−29·10	−3·11	−58·89	−44·11	+24·36
February	+47·25	−44·64	+79·42	−29·63	−3·65	−59·46	−44·54	+23·11
March	+47·11	−45·92	+78·85	−30·09	−4·19	−60·09	−44·77	+22·62
April	+46·95	−47·20	+78·47	−30·50	−4·69	−60·69	−44·63	+22·14
May	+46·90	−48·50	+78·20	−30·85	−5·07	−61·14	−44·27	+21·17
June	+46·85	−50·04	+77·89	−31·14	−5·42	−61·66	−44·18	+20·45
July	+46·84	−51·63	+77·69	−31·54	−5·84	−62·26	−44·41	+19·94
August	+46·76	−53·42	+77·61	−31·92	−6·19	−62·77	−44·70	+19·50
September	+46·51	−55·34	+77·31	−32·21	−6·47	−63·17	−45·08	+19·05
October	+46·34	−57·47	+76·74	−32·66	−6·87	−63·72	−45·57	+18·69
November	+46·30	−59·67	+76·25	−33·01	−7·19	−64·15	−45·86	+18·31
December	+46·27	−62·06	+75·54	−33·40	−7·55	−64·60	−46·11	+17·84
1952								
January	+46·23	−64·48	+74·76	−33·81	−7·89	−65·01	−46·21	+17·47
February	+46·16	−67·06	+73·97	−34·28	−8·22	−65·40	−46·25	+17·09
March	+46·05	−69·57	+75·78	−34·78	−8·60	−65·79	−46·11	+16·72
April	+46·00	−72·06	+75·40	−35·23	−8·97	−66·15	−45·90	+16·42
May	+45·92	−74·43	+72·62	−35·67	−9·28	−66·61	−45·70	+16·15
June	+45·50	−76·97	+72·31	−36·21	−9·73	−67·18	−45·62	+15·53
Criterion	**0·08**	**0·10**	**0·53**	**0·06**	**0·07**	**0·08**	**0·17**	**0·15**

Observations made over many years have revealed that the Sun, Moon and planets are departing from their predicted positions by amounts that are proportional to their mean motions (**14**). This is attributed to an error in the time scale against which the observations have been recorded: the observed secular accelerations of the Sun, Mercury and Venus are thus associated with a progressive retardation in the rate of rotation of the Earth, amounting to an increase of one millisecond per century in the length of the day. It now appears that this may be satisfactorily accounted for by tidal friction. In addition there are fluctuations in the observed motions of the Sun, Moon, Mercury and Venus,

arising from variations in the rotation of the Earth. These variations have been interpreted as occasional sudden changes of several milliseconds in the length of the day, but may alternatively be regarded as the cumulative effects of continual random changes.

It is clearly desirable that the motions of the celestial bodies should be referred to a time system unaffected by irregularities in the rate of rotation of the Earth. Such a system is defined by the period of the revolution of the Earth in its orbit round the Sun, making the sidereal year the fundamental unit. At the Conference on the Fundamental Constants of Astronomy (Paris, 1950) it was recommended that the unit adopted should be the sidereal year at 1900·0 (**15**, **16**). The time reckoned in terms of this unit is known as Ephemeris Time and the difference between mean solar time and Ephemeris Time is given by the following expression:—

$$+24^{s}\cdot349+72^{s}\cdot3165T+29^{s}\cdot949T^{2}+1^{s}\cdot821B,$$

where T is measured in Julian centuries from 1900 January 0, Greenwich Mean Noon, and B is the fluctuation in the Moon's longitude, that is, the difference between the observed longitude and the amended tabular longitude (**14**). The previously accepted unit of time, the mean solar second, which is now known to be variable, may be converted to an invariable unit, the second of Ephemeris Time, by use of the same formula.

The second is one of the three fundamental physical units, and thus enters into many scientific measurements besides those of astronomy. In the majority of cases, the errors arising from the use of the variable mean solar second are negligible, but clearly an invariable unit of time should be employed. The second of Ephemeris Time is specifically intended to provide a unit of this kind, but its widespread use is dependent upon the development of satisfactory methods for its current determination. This requires regular accurate observations of the position of the Moon for comparison with the tabular places. Work now in progress at the U.S. Naval Observatory suggests that observations of sufficient accuracy may be secured by the employment of a new photographic technique. These developments may make it possible for a time-keeping observatory to establish a uniform time system, and thus provide a satisfactory basis for the assessment of clock performance and for frequency standardization.

3. *Assessment of short-term performance.*—It will be evident that the day-to-day stability of quartz clocks over periods of a month or so may be satisfactorily assessed from the clock intercomparisons alone without reference to the astronomical observations. As may be expected, the clocks which have been selected in virtue of their high long-term stability also exhibit a high standard of uniformity from day to day. An example is afforded by the relative daily rate of the clocks 9A and 9C, which is plotted in Fig. 5. Throughout the year 1951 the variation in rate was smooth, and the range did not exceed 0·5 millisecond per day. It may be remarked that the drifts of the clocks are sufficiently similar for the relative rate to be plotted direct, without removal of an ephemeris.

The most satisfactory indication of day-to-day performance is provided by an examination of the second differences of the daily values of clock rate. The mean absolute value of the second difference affords a quantitative criterion of the suitability of the clock for day-to-day extrapolation and interpolation. It has recently been pointed out (**17**) that there is some correlation between the second

differences on successive days. It should not be assumed, however, that this necessarily impairs their utility as a measure of short-term performance, since it has been shown (**6**) that they may be grouped into three sub-sequences, each of which is entirely free from internal correlation. The re-combination of the sub-sequences does not invalidate the result. In practice, it is unnecessary to separate the sub-sequences and re-combine when forming the mean absolute second difference.

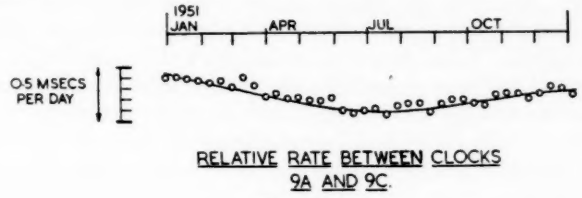

FIG. 5.

Clock intercomparisons of high accuracy are obtainable from the daily readings of the rotary beat counters, which give a continuous indication of the time difference between any two clocks in units of one microsecond. The calculation of the second differences of daily rates is readily performed by means of a National Accounting Machine. These daily values are plotted to give a convenient picture of the day-to-day stability of the clocks. The mean absolute value of the second difference of daily rate provides a useful criterion of short-term performance. The intercomparisons between the two clocks E5 and E6 throughout the year 1951 gave a mean absolute daily second difference of relative rate amounting to 16 microseconds per day. Since the two clocks were of comparable short-term stability, the criterion value was approximately 11 (in the same units) for each clock. The intercomparisons between C5 and C6 for a period of some four or five months after the modification of the C6 maintaining amplifier gave a mean absolute daily second difference of 19 microseconds per day, corresponding to a criterion value of 13 for each clock, if the stabilities were equal. It is clear that a short-period criterion value of 10 to 15 is attained by modern quartz clocks. While this accuracy cannot be fully utilized in Time Service operational work, the precision of the daily intercomparisons provides a valuable indication of any slight deterioration in the stability of the clocks which might, by its cumulative effect, cause a significant deterioration in the long-term performance. It may sometimes be possible to associate an enhancement of these small variations with particular physical causes, such as alterations in the values of certain components, thus providing information of great value in the modification and design of quartz-crystal clocks.

A graph of the direct intercomparisons of E5/E6 and C6/E6 with a parabolic ephemeris removed in each case is given in Fig. 6. It will be noted that the deviations do not exceed 0·1 millisecond in the course of a month.

4. *Present developments.*—The high stability of the performance attained by modern clocks leads naturally to the consideration of the possibility of further progress. It has been suggested that the ultimate limit of accuracy is imposed by the crystal itself, rather than by the ancillary circuits. If this were indeed the case, consideration would have to be given to the development of standards utilizing some control element other than a quartz crystal. Tests on the ancillary

circuits have shown, however, that in many cases the crystal is not operating under the best conditions. It has in fact been possible, by improved design of the associated circuits, to obtain a higher standard of stability from crystals of conventional types.

In particular, attention was directed to the reduction of the amplitude of oscillation of the crystal. This cannot be achieved in the conventional bridge-drive circuit, where the amplitude of oscillation is controlled directly by a tungsten filament lamp. As a first step, the tungsten filament lamp was replaced by an indirectly heated thermistor, for which the heating current was derived from a secondary amplifier. This gave some reason to believe that higher short-term stability would result from a lower amplitude of oscillation. In the most recent drive-circuit, the bridge has been replaced by a circuit which is electrically equivalent, and in which the crystal current is less than 30 micro-amperes. The maintaining amplifier uses no transformers, as these were thought to be a source of instability in the circuits previously employed. The achievement of higher precision in the crystal oscillator makes it advantageous to control the temperature of the crystal within closer limits and it is estimated that a new control circuit and oven divide ambient changes by a factor of 5000

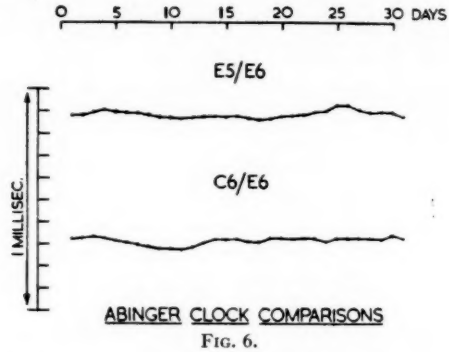

FIG. 6.

Reliability of operation over long periods is of paramount importance, since the clock may often suffer greater disturbance from a failure of temperature control than from a short interruption of oscillation. The previous designs used the 50 c/s mains supply both for the control bridge and for the heater windings. In the new experimental oscillators the current is supplied by a battery-operated 325 c/s oscillator which is amplitude-stabilized. The oscillator valves are in parallel, and removal of either valve has negligible effect on the frequency and amplitude. The controlling amplifier, which employs duplicated hard valves in place of thyratrons, has been designed in such a way that the valves may be replaced without altering the control-temperature by more than half a millidegree. The first oscillator incorporating these refinements employed a GT-cut plate crystal, and was put into operation in 1951 November. A second oscillator, incorporating minor improvements, was installed in 1952 July. Results obtained so far have been encouraging.

5. *Conclusions.*—It is now nearly ten years since the quartz clock, by virtue of its superiority as a long-term standard, displaced the pendulum clock in the fundamental work of the Greenwich Time Service. Almost at once it became

apparent that among the advantages of the new type of clock were a considerably higher standard of short-period stability and far better facilities for accurate intercomparisons. The purpose of this account is to show the further advances which have subsequently been made in the quality of the clocks themselves and in the technique of using them in the operation of a national time service.

The past years have afforded an unrivalled opportunity for the gaining of experience and the investigation of the problems concerning quartz clocks. In those observatories which have not yet acquired quartz clocks comparable in stability with those available to the Greenwich Time Service, there is perhaps inadequate realization of the standards now attainable, combined with an instinctive reluctance to forsake the relatively simple pendulum clock in favour of the admittedly more complex quartz-controlled oscillator. The pendulum clock has the valuable property of being capable of prolonged uninterrupted runs, and it has been claimed that this feature is sufficient to justify the retention of pendulum clocks as fundamental standards, and that the long-term performance of the two types of clock is otherwise comparable (**18**).

Considering first the question of length of uninterrupted run, it has been found that modern quartz clocks have attained a high standard of reliability. Given continuity of power supplies, the period of uninterrupted run may easily extend over several years. Among the clocks mentioned in this account, F1 has run continuously since its installation in 1947 July, 9C since 1947 October, and 9A since 1948 January. The clocks at Abinger have not run for such long periods, as the installation has been modified in recent years. It is worthy of note, however, that in the six years since the Abinger clocks were transferred to battery operation there has been only one failure of a crystal oscillator.

As regards performance, the superiority of the quartz clock is firmly established. It is now some years since pendulum clocks were employed as standards in the Greenwich Time Service; but up-to-date information on the performance of high-precision pendulum clocks at Paris is published regularly in the *Bulletin Horaire*. Mean daily rates are quoted for each month, and these may be compared with the corresponding figures for quartz clocks given in Table I. It is also possible to calculate the mean absolute second differences of the monthly rates, thus giving a numerical criterion of performance which may be compared with the values deduced for quartz clocks. During the two years 1950–1951 one of the Paris pendulum clocks, Leroy 1372, attained a standard which can have been only rarely equalled by any other pendulum clock. The mean absolute monthly second difference was found to be 3·4 milliseconds per day. An examination of the other pendulum clocks shows that a value of 10·0 would represent a good average performance, while a criterion of 50·0 or over would be classed as poor. Results are also quoted for the quartz clocks at Paris. The best of these has a criterion of 8·5, whereas the other two have values of about 20·0. It will be evident that, despite recent claims to the contrary (**19**), even the best pendulum clock does not represent a serious challenge to the typical modern quartz clock with a criterion of 0·1 millisecond per day per month per month.

The high short-term stability of a quartz clock has permitted a new standard of accuracy in the control of radio time signals, and the application of frequency measuring technique has made it possible to carry out clock comparisons with a speed and precision hitherto unattainable. But it is perhaps in the realm of

long-term performance that the greatest advances have been made. Quartz clocks were first introduced in the Greenwich Time Service as long-term standards, and it is due to the high stability of the present clocks over periods of two or three years that it has become possible to determine the nature of the annual fluctuation in the rotation of the Earth. In current practice the annual fluctuation is determined by referring the astronomical observations to each of a number of selected clocks. The various solutions show a high degree of agreement, and the resulting mean value can be adopted with confidence.

6. *Acknowledgments.*—The various investigations and analyses described were undertaken in the course of normal routine work by the Computing Section of the Time Department, under the direction of Mr R. H. Tucker, who also assisted in the preparation of the paper. Information relating to the quartz clocks at other establishments is regularly supplied to the Royal Observatory and is of great value in the operation of the National Time Service. The author wishes to take this opportunity of expressing his appreciation of the continued cooperation of the Post Office Radio Experimental and Development Branch Laboratories and of the National Physical Laboratory.

Royal Greenwich Observatory,
Abinger Common,
Dorking, Surrey :
1952 *October* 9.

References

(1) L. Essen, *Proc. Phys. Soc.*, **50,** 413, 1938.
(2) L. A. Meacham, *Bell System Technical Journal*, **17,** 574, 1938.
(3) L. B. Turner, *Journal I.E.E.*, **81,** 399, 1937.
(4) R. L. Miller, *Proc. I.R.E.*, **27,** 446, 1939.
(5) C. F. Booth and F. J. M. Laver, *Journal I.E.E.*, **93,** Part III, 223, 1946.
(6) W. M. H. Greaves and L. S. T. Symms, *M.N.*, **103,** 196, 1943.
(7) H. M. Smith, *Annales Françaises de Chronométrie*, 2e Série, **3,** 227, 1949.
(8) H. M. Smith, *British Science News*, **2,** 169, 1949.
(9) H. M. Smith, *British Science News*, **2,** 207, 1949.
(10) H. B. Law, *Journal I.E.E.*, **94,** Part III, 38, 1947.
(11) W. A. Marrison, *Bell System Technical Journal*, **27,** 510, 1948.
(12) H. M. Smith, *Proceedings I.E.E.*, **98,** Part II, 143, 1951.
(13) H. M. Smith, *I.E.E. Monograph*, No. 39.
(14) H. Spencer Jones, *M.N.*, **99,** 541, 1939.
(15) *Bulletin Astronomique*, tome XV, fascicule 4, 291, 1950.
(16) D. H. Sadler, *M.N.*, **111,** 624, 1951.
(17) N. Stoyko, *Annales Françaises de Chronométrie*, 2me Série, **3,** 139, 1949.
(18) R. Forga, *l'Astronomie*, 65e année, 112, 1951.
(19) A. Danjon, *Annales Françaises de Chronométrie*, 2me Série, **3,** 365, 1949.

ON THE MASSES OF SATURN'S SATELLITES

Sir Harold Jeffreys, F.R.S.

(Received 1953 January 12)

Summary

The principal determinations of the masses of Saturn's satellites are rediscussed, with special attention to the standard errors. A new determination of the mass of Titan is made from the motion of the orbital plane of Iapetus, and an attempt is made to estimate the errors remaining in Woltjer's solution from Hyperion. The values resulting are as follows, expressed as fractions of the mass of Saturn:

Mimas $(6{\cdot}69\pm0{\cdot}20)10^{-8}$; Enceladus $(1{\cdot}27\pm0{\cdot}53)10^{-7}$;
Tethys $(1{\cdot}141\pm0{\cdot}030)10^{-6}$; Dione $(1{\cdot}825\pm0{\cdot}061)10^{-6}$;
Rhea $(0{\cdot}4\pm3{\cdot}8)10^{-6}$; Titan $(2{\cdot}412\pm0{\cdot}018)10^{-4}$;

and Struve's constants for Saturn are

$$k/a_0^2=0{\cdot}024\ 305\pm0{\cdot}000\ 048,$$
$$l/a_0^4=(11{\cdot}0\pm2{\cdot}0)10^{-4}.$$

---

1. The mass of Titan was determined almost simultaneously by S. Newcomb from the motion of the apse of Hyperion and by H. Struve from a comparison of the motions of the orbital plane of Iapetus and the apse of Titan. The results differed by a factor of nearly 3, but successive corrections to the theory of Hyperion have produced a rough agreement. There seems, however, to have been no new solution from Iapetus. Struve's discussion* is statistically very unsatisfactory. It is repeated with little alteration by Tisserand.† Using the same data, I applied the method of least squares, supposing that standard errors in i and θ (longitude of the node) were in the ratio sin i to 1, and giving equal weight to all data. The result was (the mass of Saturn being taken as 1)

$$m_{\mathrm{Ti}}=(2{\cdot}324\pm0{\cdot}075)10^{-4},$$
$$1/m_{\mathrm{Ti}}=4\,308\pm138\ \text{(s.e.)}.$$

Struve's value was given as $4\,678\pm350$. The standard error suggests that with modern data the method might yield an uncertainty under 1 per cent.

J. Woltjer‡ gives seven determinations from Hyperion, ranging from 3 767 to 4 172, and adopts the value 4 033.

H. Struve§ discussed the observations of all the satellites and made important contributions to the theory. G. Struve‖ rediscussed the observations, with others up to 1928, and derived new masses. Estimates of uncertainty are given

* H. Struve, *Beobachtungen d. Saturnstrabanten, Supp.* 1 *aux Observations de Poulkova*, 1888.
† F. Tisserand, *Théorie de Mécanique Céleste*, **4**, 98–102, 1896.
‡ J. Woltjer, *Ann. d. Sternwarte Leiden*, **16**, Part 3, 1928.
§ H. Struve, *Publ. Obs. Centr. Nicholas*, **11**, 1898.
‖ G. Struve, *Veröff. Berlin-Babelsberg*, **6**, Hefte 1–5, 1933.

for the equations of condition, but the final masses and their uncertainties are based on judgment and not on least squares solutions. Accordingly it cannot be said that their uncertainties are known, which is unfortunate in view of their importance for an understanding of the constitutions of these bodies. In the course of the present work it was found that G. Struve's uncertainties ranged from half to twice the s andard errors.

2. The following analysis is mainly based on G. Struve's data. Some of the equations of condition were rediscussed but no important discrepancy was found. Uncertainties have been converted to standard errors.

The motions of the nodes and apses derived from the secular terms in the disturbing functions due to the ellipticity of Saturn and the attractions of the satellites are denoted by $\pm\beta n$; they are theoretically equal and opposite, to the accuracy available. Approximate commensurability of the mean motions produces systematic changes of the secular motions for the pairs Enceladus–Dione and Titan–Hyperion. The solar effect has been subtracted from all observed values. The theoretical values as functions of the masses are given by G. Struve (Heft 4, p. 53) for the first six satellites. A discussion of the observations of Titan and Iapetus is given in the posthumous Heft 5.

For the pairs Mimas–Tethys and Enceladus–Dione, though the masses are small, good determinations are possible because fairly large perturbations are produced by approximate commensurability of the mean motions.

The mean motions in degrees per day given by G. Struve will be adopted (Table I). The value for Hyperion is Woltjer's.

TABLE I

Mimas	381·994 442 ± 0·000 013
Enceladus	262·731 940 5 ± 0·000 002 2
Tethys	190·697 950 ± 0·000 005
Dione	131·534 972 9 ± 0·000 001 2
Rhea	79·690 088 1 ± 0·000 000 7
Titan	22·577 015 08 ± 0·000 000 40
Hyperion	16·919 989 6 ± 0·000 002 6
Iapetus	4·537 995 36 ± 0·000 000 40

The observed values of βn, in degrees per year, are in Table II.

TABLE II

	ϖ	θ	
Mimas	365·60 ± 0·10	365·23 ± 0·10	(Heft 4, pp. 18, 47)
Enceladus	...	...	
Tethys	...	72·227 ± 0·031	(p. 42)
Dione	30·75 ± 0·21	...	(p. 33)
Rhea	...	10·20 ± 0·04	(p. 40)
Titan	0·501 2 ± 0·003 5	0·492 ± 0·029	

The mean motion of the apse of Titan is given (Heft 5, pp. 12–13) as $31'{\cdot}41 \pm 0'{\cdot}21$ per year; from this must be subtracted the solar part $1'{\cdot}34$ (H. Struve, p. 172), leaving the value given. The value for the node is a rough determination from the data in G. Struve's table (Heft 5, p. 11). The values for Iapetus need further discussion.

I consider first the periodic disturbances.

2.1. *Mimas and Tethys.*—The theory is outlined by Tisserand (pp. 129–138). G. Struve (Heft 4, p. 10) writes the perturbations of mean longitude of Mimas in the form

$$\Delta l = A \sin \mu(\tau - \tau_0) + B \sin 3\mu(\tau - \tau_0)$$

and determines

$$A = -44^\circ{\cdot}390 \pm 0^\circ{\cdot}122,$$

$$\mu = (5^\circ{\cdot}0864 \pm 0^\circ{\cdot}0085)/\text{year},$$

$$\tau_0 = 1\,866{\cdot}27 \pm 0{\cdot}06.$$

For Tethys A is replaced by A_1 and the same values of μ, τ_0 are adopted; then

$$A_1 = 2^\circ{\cdot}065 \pm 0^\circ{\cdot}028.$$

The ratio $m_{\text{Mi}}/m_{\text{Te}}$ then follows from

$$\frac{a_1 m_{\text{Mi}}}{a m_{\text{Te}}} = 2\frac{A_1}{A} = 0{\cdot}093\,04.$$

The value of m_{Te} depends on the use of elliptic functions, but substantially

$$m_{\text{Te}} \propto \frac{1}{\mu^2 \gamma \gamma_1},$$

where γ, γ_1 are the inclinations of the two orbital planes to Saturn's equator. Allowing for the uncertainties of these and of μ gives a proportional standard error of $\pm 2 \times 0{\cdot}0017 \pm 0{\cdot}025 \pm 0{\cdot}009 = \pm 0{\cdot}026$. Then

$$m_{\text{Te}} = \frac{1}{876\,400}(1 \pm 0{\cdot}026) = 1{\cdot}141 \times 10^{-6}(1 \pm 0{\cdot}026).$$

The relative uncertainty of m_{Mi} is 0·014 from that of A_1; that of A is unimportant, but those of μ, γ, γ_1 enter again; then

$$m_{\text{Mi}} = \frac{1}{14\,960\,000}(1 \pm 0{\cdot}026 \pm 0{\cdot}014)$$

$$= 6{\cdot}69 \times 10^{-8}(1 \pm 0{\cdot}030).$$

G. Struve gives 837 000 and 915 000 as extreme values of $1/m_{\text{Te}}$, corresponding to about 1·5 times the standard error. On the other hand, he gives the uncertainty of the mass of Mimas as 5–10 per cent, the latter being too high on a similar scale.

2.2. *Enceladus and Dione.*—The largeness of the perturbations in this case also depends on the fact that the mean motions are nearly as 2 to 1, but the theory is much more difficult. In the case of Mimas and Tethys the argument of the libration is $4l_1 - 2l - \theta - \theta_1$, and thus depends on a product of the inclinations and not on the eccentricities. For Enceladus and Dione it does depend on the eccentricities. H. Struve used a linearized theory, which is used in part by G. Struve. Woltjer* pointed out that higher terms need attention, and G. Struve quotes his solution also. It allows, however, only for perturbations of Enceladus by Dione, and a somewhat fuller theory is needed. The following method is that of H. Struve extended to allow for certain second-order terms. Letters with suffix 1 refer to Dione, without them to Enceladus.

* J. Woltjer, *Bull. Astr. Inst. Neth.*, **1**, 23–24 and 36, 1922.

The rates of change of the relevant orbital elements are given by (H. Struve, 1898, p. 178)

$$\frac{dn}{dt} = 3m_1n^2\{Ae\sin(2l_1-l-\varpi) - Be_1\sin(2l_1-l-\varpi_1)\}, \tag{1}$$

$$\frac{de}{dt} = m_1nA\sin(2l_1-l-\varpi), \tag{2}$$

$$\frac{d\varpi}{dt} = \beta n - m_1n\frac{A}{e}\cos(2l_1-l-\varpi), \tag{3}$$

$$\frac{dn_1}{dt} = -6mn_1^2\{A_1e\sin(2l_1-l-\varpi) - B_1e_1\sin(2l_1-l-\varpi_1)\}, \tag{4}$$

$$\frac{de_1}{dt} = -mn_1B_1\sin(2l_1-l-\varpi_1), \tag{5}$$

$$\frac{d\varpi_1}{dt} = \beta_1n_1 + mn_1\frac{B_1}{e_1}\cos(2l_1-l-\varpi_1), \tag{6}$$

$$A = 0{\cdot}7532, \quad B = 0{\cdot}2714, \quad A_1 = 1{\cdot}1943, \quad B_1 = 0{\cdot}4348, \quad AB_1 - A_1B = 0{\cdot}0034. \tag{7}$$

Put

$$V_0 = (2n_1-n)t, \qquad V = 2l_1-l = V_0+V', \tag{8}$$

$$h = e\sin(\varpi-V_0), \quad k = e\cos(\varpi-V_0), \quad h = e_1\sin(\varpi_1-V_0), \quad k = e_1\cos(\varpi_1-V_0), \tag{9}$$

$$\nu = \beta n - (2n_1-n), \quad \nu_1 = 2n_1-n-\beta_1n_1. \tag{10}$$

Then

$$\frac{dh}{dt} = \nu k - m_1nA\cos V', \tag{11}$$

$$\frac{dk}{dt} = -\nu h + m_1nA\sin V', \tag{12}$$

$$\frac{dh_1}{dt} = -\nu_1k_1 + mn_1B_1\cos V', \tag{13}$$

$$\frac{dk_1}{dt} = \nu_1h_1 - mn_1B_1\sin V', \tag{14}$$

$$\frac{d^2l}{dt^2} = \frac{dn}{dt} = 3m_1n^2(-Ah\cos V' + Ak\sin V' + Bh_1\cos V' - Bk_1\sin V'), \tag{15}$$

$$\frac{d^2l_1}{dt^2} = \frac{dn_1}{dt} = -6mn_1^2(-A_1h\cos V' + A_1k\sin V' + B_1h_1\cos V' - B_1k_1\sin V'), \tag{16}$$

$$\frac{d^2V'}{dt^2} = P(h\cos V' - k\sin V') - Q(h_1\cos V' - k_1\sin V'), \tag{17}$$

where

$$P = 12mn_1^2A_1 + 3m_1n^2A, \tag{18}$$

$$Q = 12mn_1^2B_1 + 3m_1n^2B \doteqdot BP/A. \tag{19}$$

For any constant α, γ

$$V' = \alpha + \gamma t$$

leads to an exact solution of (11) to (14) and (17). This corresponds to the addition of constants to n, n_1 and the epochs. But it also indicates that the period

equation contains γ^2 as a factor, and that we can choose the constants so that V'' has no constant or secular part. This identifies n, n_1 with the observed mean motions.

For V' small, if we neglect V'^2, there is a constant solution

$$h = h_1 = 0, \quad V' = 0,$$

$$k = k_0 = \frac{m_1 n A}{\nu}, \quad k_1 = k_{10} = -\frac{m n_1 B_1}{\nu_1}. \tag{20}$$

Accents denoting departures from this constant solution, we put

$$h = c \sin \gamma t, \quad k' = \frac{\gamma}{\nu} c \cos \gamma t, \quad h_1 = c_1 \sin \gamma t, \quad k_1' = -\frac{\gamma}{\nu_1} c_1 \cos \gamma t, \quad V' = \lambda \sin \gamma t \tag{21}$$

and retain the first powers of c, c_1, λ. Then, cancelling a factor $\gamma^2\lambda$, we have the period equation

$$\frac{P k_0}{\gamma^2 - \nu^2} - \frac{Q k_{10}}{\gamma^2 - \nu_1^2} = 1 \tag{22}$$

with

$$c = -\frac{m_1 n A \nu \lambda}{\gamma^2 - \nu^2}, \quad c_1 = \frac{m n_1 B_1 \nu_1 \lambda}{\gamma^2 - \nu_1^2}.$$

To the second order in k_0, k_{10} the roots are given by

$$\gamma^2 - \nu^2 = P k_0 \left(1 - \frac{Q k_{10}}{\nu^2 - \nu_1^2}\right), \tag{23}$$

$$\gamma_1^2 - \nu_1^2 = -Q k_{10} \left(1 - \frac{P k_0}{\nu^2 - \nu_1^2}\right), \tag{24}$$

and

$$\Delta l = -3 m_1 n^2 \left\{\left(\frac{A k_0}{\gamma^2 - \nu^2} - \frac{B k_{10}}{\gamma^2 - \nu_1^2}\right) \lambda \sin \gamma t + \left(\frac{A k_0}{\gamma_1^2 - \nu^2} - \frac{B k_{10}}{\gamma_1^2 - \nu_1^2}\right) \lambda_1 \sin \gamma_1 t\right\}, \tag{25}$$

$$\Delta l_1 = 6 m n_1^2 \left\{\left(\frac{A_1 k_0}{\gamma^2 - \nu^2} - \frac{B_1 k_{10}}{\gamma^2 - \nu_1^2}\right) \lambda \sin \gamma t + \left(\frac{A_1 k_0}{\gamma_1^2 - \nu^2} - \frac{B_1 k_{10}}{\gamma_1^2 - \nu_1^2}\right) \lambda_1 \sin \gamma_1 t\right\},$$

$$\doteq -\frac{2 m n_1^2}{m_1 n^2} \frac{A_1}{A} \Delta l = -\frac{1}{2} \frac{m}{m_1} \frac{A_1}{A} \Delta l. \tag{26}$$

k_0 is small and k_{10} still smaller, and if we neglect m the solution reduces to that of Woltjer. Note that Δl, Δl_1 are not small compared with λ, λ_1, since $\gamma^2 - \nu^2$ and $\gamma_1^2 - \nu_1^2$ contain the small factors m, m_1.

G. Struve gives l, e and ϖ for mean dates, each summarizing observations over about a year. With e and ϖ he gives also $2l_1 - l$, which never differs from ϖ by more than a few degrees. Thus k from (9) is practically the recorded e, and the mean value of the latter is an estimate of k_0. Thus k_0 is the forced eccentricity of Enceladus. The variations of e and of $e \sin (2l_1 - l - \varpi)$ are however too small to give useful determinations of the periodic parts of h and k. The values of e give

$$k_0 = 0{\cdot}004\,48 \pm 0{\cdot}000\,21. \tag{27}$$

G. Struve's value (p. 33) is $0{\cdot}004\,44 \pm 0{\cdot}000\,09$ (p.e.) but the uncertainty seems to be a mistake.

For Dione the mean eccentricity given by Struve's table (p. 34) is

$$\bar{e}_1 = 0{\cdot}002\,11 \pm 0{\cdot}000\,09. \tag{28}$$

Struve gives 0·002 21 without an uncertainty. (Some of these small changes are due to differences in the allocation of weights where two sets of observations correspond to one year. I have checked most of the summary values by means of χ^2, and took the larger uncertainty in cases of disagreement, which seldom occurred.) As for Enceladus the periodic variations h_1', k_1' are too small to be detected from the variations of e and ϖ.

The periodic parts of l are found to be (G. Struve, p. 23)

$$\Delta l = (14'\cdot39 \pm 1'\cdot20)\sin\{63^\circ\cdot75 \pm 5^\circ\cdot28 + (32^\circ\cdot51 \pm 0^\circ\cdot30)t\} + (14'\cdot06 \pm 1'\cdot19)\sin\{117^\circ\cdot28 \pm 5^\circ\cdot47 + (93^\circ\cdot14 \pm 0^\circ\cdot40)t\}, \quad (29)$$

and direct comparison of this with l_1 gives

$$-\frac{\Delta l_1}{\Delta l} = x = +0\cdot065 \pm 0\cdot025. \quad (30)$$

The speeds of the harmonic terms are estimates of γ, γ_1, but for their interpretation we need also preliminary estimates of βn, $\beta_1 n_1$. The free eccentricity of Dione ($\bar{e}_1$) leads observationally to

$$d\varpi_1/dt = 30^\circ\cdot75 \pm 0^\circ\cdot21; \quad (31)$$

there is no direct evidence at all for Enceladus. H. and G. Struve proceeded by interpolation from the values for Mimas and Tethys. I extend their method by fitting a form $Cn^{7/3} + Dn^{-1}$ to all the observed values out to Rhea, giving equal weights to each. This amounts roughly to regarding the secular motions as mainly due to the ellipticity of Saturn and perturbations by Titan, with a small part due to roughness of the Titan term and to the effects of the other satellites, which is taken as random. The object is only to get preliminary estimates of βn, $\beta_1 n_1$, which can be improved later. The result is, with βn in degrees per year, n in degrees per day,

$$C = 3\cdot450\,08 \times 10^{-4}, \quad D = +53,$$

with the following results.

		O–C
Mimas	$365\cdot26 + 0\cdot14 = 365\cdot40$	$-0\cdot17, +0\cdot20$
Enceladus	$152\cdot57 + 0\cdot20 = 152\cdot77$	
Tethys	$72\cdot23 + 0\cdot28 = 72\cdot51$	$-0\cdot28$
Dione	$30\cdot35 + 0\cdot40 = 30\cdot75$	$0\cdot00$
Rhea	$9\cdot43 + 0\cdot67 = 10\cdot10$	$+0\cdot10$

The estimated standard error of one value is 0·23. This is more than the apparent uncertainties of the observed values, and it will be worth while to attempt greater detail. It is clear, however, that the two observed values for Mimas differ by 2·5 times the standard error of their difference, and that the interpolation gives βn for Enceladus with as much accuracy as the speeds of the two harmonic terms already possess. For Dione the interpolate happens to agree exactly with the observed value. I take

$$\beta n = 152^\circ\cdot77/1^y + \delta(\beta n), \quad (32)$$

$$\beta_1 n_1 = 30^\circ\cdot75/1^y + \delta(\beta_1 n_1), \quad (33)$$

where the corrections are unlikely to exceed $0^\circ\cdot3/1^y$. Then

$$2n_1 - n = 123^\circ\cdot44/1^y, \quad \nu = 29^\circ\cdot33 - \delta(\beta n), \quad \nu_1 = 92^\circ\cdot69 + \delta(\beta_1 n_1). \quad (34)$$

I take also

$$m = 1{\cdot}5 \times 10^{-7}(1+\mu), \quad m_1 = 1{\cdot}8 \times 10^{-6}(1+\mu_1). \tag{35}$$

Then

$$k_0 = 0{\cdot}004\,46\,(1+\mu_1)\,(1+\tfrac{1}{29}\delta(\beta n)), \tag{36}$$

$$k_{10} = -0{\cdot}000\,026\,0\,(1+\mu)\,(1-\tfrac{1}{93}\delta(\beta_1 n_1)), \tag{37}$$

$$\frac{Pk_0}{n^2} = (2{\cdot}056 + 0{\cdot}238\mu + 3{\cdot}870\,\mu_1)\,10^{-8}\,(1+\tfrac{1}{29}\delta(\beta n)), \tag{38}$$

$$\frac{Qk_{10}}{n^2} = -(4{\cdot}32 + 4{\cdot}82\,\mu + 3{\cdot}81\,\mu_1)\,10^{-11}\,(1-\tfrac{1}{93}\delta(\beta_1 n_1)), \tag{39}$$

whence, with sufficient accuracy,

$$\frac{\gamma}{\nu} = (1{\cdot}105\,7 + 0{\cdot}010\,6\,\mu + 0{\cdot}172\,3\,\mu_1)\,(1-\tfrac{1}{290}\delta(\beta n)), \tag{40}$$

$$\gamma = 32{\cdot}43 + 0{\cdot}34\,\mu + 5{\cdot}59\,\mu_1 - \delta(\beta n), \tag{41}$$

and

$$0{\cdot}34\,\mu + 5{\cdot}59\,\mu_1 = +0{\cdot}08 \pm 0{\cdot}30 + \delta(\beta n). \tag{42}$$

From k_0

$$0{\cdot}004\,46\,(\mu_1 + \tfrac{1}{29}\delta(\beta n)) = +0{\cdot}000\,02 \pm 0{\cdot}000\,21. \tag{43}$$

Also

$$(\gamma_1^2 - \nu_1^2)/\nu_1^2 \simeq 10^{-5} \tag{44}$$

and is negligible. The ratio $\Delta l_1/\Delta l$ leads to

$$0{\cdot}066\,1\,\mu - 0{\cdot}066\,1\,\mu_1 = -0{\cdot}001\,1 \pm 0{\cdot}025. \tag{45}$$

The mean eccentricity of Dione is $c_1 = \bar{e}_1$; the amplitude of the mode with speed γ_1 for Enceladus is Struve's q,

$$q = 3B\,m_1\,c_1\left(\frac{n}{\nu_1}\right)^2 = 5\,459'\,c_1\,(1+\mu_1) = 14'{\cdot}06 \pm 1'{\cdot}20. \tag{46}$$

In $c_1\mu_1$ we take c_1 from (28); then

$$c_1 + 0{\cdot}002\,1\,\mu_1 = 0{\cdot}002\,58 \pm 0{\cdot}000\,22. \tag{47}$$

$\delta(\beta n)$ and $\delta(\beta_1 n_1)$ have been neglected where their effects are negligible in comparison with the uncertainties already present. On account of (44) the "periodic" term in the disturbing function due to Enceladus produces a negligible effect in the motion of the apse of Dione.

Then the equations of condition for μ, μ_1, c_1 are (42), (43), (28), (45), (47). The solution is

$$\mu = -0{\cdot}027 + 0{\cdot}37\,\delta(\beta n) \pm 0{\cdot}35, \tag{48}$$

$$\mu_1 = +0{\cdot}030 + 0{\cdot}047\,\delta(\beta n) \pm 0{\cdot}034, \tag{49}$$

$$c_1 = 0{\cdot}002\,19 - 0{\cdot}000\,02\,\delta(\beta n) \pm 0{\cdot}000\,08. \tag{50}$$

Comparing with the equations of condition, and ignoring $\delta(\beta n)$, we get $\chi^2 = 3{\cdot}2$ on 2 degrees of freedom. Then

$$m = 1{\cdot}46 \times 10^{-7}\,\{1 + 0{\cdot}38\,\delta(\beta n) \pm 0{\cdot}36\}, \tag{51}$$

$$m_1 = 1{\cdot}854 \times 10^{-6}\,\{1 + 0{\cdot}046\,\delta(\beta n) \pm 0{\cdot}033\}, \tag{52}$$

$$1/m \doteqdot 6\,850\,000, \quad 1/m_1 \doteqdot 539\,400. \tag{53}$$

G. Struve gives 6 622 000 and 541 300, with uncertainties of 20 and 5 per cent. That for Enceladus is too low; that for Dione is about 1·5 times the standard error.

The solution is not quite so satisfactory as it appears, for a reason given by G. Struve (p. 37). The argument of the γ_1 term is essentially $2l_1 - l - \varpi_1$, and the constant term in this is (with probable errors)

$$-173^\circ{\cdot}4 \pm 3^\circ{\cdot}4 + (507^\circ\,44'{\cdot}0 \pm 0'{\cdot}8) - (199^\circ\,25'{\cdot}8 \pm 0'{\cdot}8) = 134^\circ{\cdot}90 \pm 3^\circ{\cdot}4.$$

The observed value is $117^\circ{\cdot}28 \pm 3^\circ{\cdot}69$, and the difference is $17^\circ{\cdot}62 \pm 5^\circ{\cdot}15$ (p.e.) or $\pm 7^\circ{\cdot}62$ (s.e.). This would have contributed 5·4 to χ^2 had it been included in the comparison, making $\chi^2 = 8{\cdot}6$ on 3 degrees of freedom. It did not arise in the actual solution because the phase affects none of the equations of condition, but there are also certain persistences of sign in the residuals that suggest unconsidered long-period perturbations. If these have been inadequately eliminated in the method of solution the uncertainties may need to be increased by a factor of 1·5 or so.

2.3. *Rhea.*—This has a forced eccentricity due to Titan, and the pericentre oscillates about that of Titan. G. Struve (Heft 1, p. 11) adopts $m_{\mathrm{Ti}} = \frac{1}{4033}$ and derives the libration constants by least squares. The residuals of e after allowance for the libration give a mean $(-2 \pm 6{\cdot}7)\,10^{-5}$; the adopted mean was 0·000 98. Taking the mean eccentricity as proportional to m_{Ti} we get

$$m_{\mathrm{Ti}} = \frac{96 \pm 6{\cdot}7}{98} \times \frac{1}{4033} = 0{\cdot}000\,242\,9\,(1 \pm 0{\cdot}070),$$

$$\frac{1}{m_{\mathrm{Ti}}} = 4\,117 \pm 288.$$

2.4. *Titan.*—For the mean motion of the node on the proper plane G. Struve adopted H. Struve's value of 0°·506/year, and he apparently did not try to improve it. For the apse and node I take the values in Table II.

An estimate of the mass of Iapetus is derived from the inclination of the orbit to Saturn's equator, namely

$$m_{\mathrm{Ja}} = 2{\cdot}46\,(1 \pm 0{\cdot}78) \times 10^{-6}.$$

The uncertainty has as usual been converted into a standard error. The result is only an indication that the mass is comparable with those of Tethys and Dione.

Examination of possible effects of Hyperion shows that its mass is less than $\frac{1}{3000}$ of that of Titan.

2.5. *Hyperion.*—The theory of this satellite is far the most complicated of all. Woltjer's treatment* is the fullest. The essential points may be indicated by comparison with the perturbations of Enceladus by Dione. The chief differences are that the mean motions are nearly in the ratio 4 to 3 instead of 2 to 1; the reaction of Hyperion on Titan can be neglected; and the mass of Titan is about 100 times that of Dione, so that approximations have to be carried to a much higher order. As the counterpart of k_{10} is negligible, that of γ_1 reduces to ν_1.

The resemblances are as follows. (1) The k_0 term in the motion of Enceladus has a much larger counterpart. This is a forced eccentricity of about 0·1 combined with a motion of pericentre $4n - 3n'$, accents referring to Titan. (2) The free oscillations about the resulting periodic orbit, in the case of Enceladus, are represented by the term in mean longitude of speed γ. When the departures from a circular orbit are described in terms of an eccentricity and a moving apse, and the free eccentricity is less than the forced one, the apse can be described as librating about a point with constant motion. Hyperion shows a similar feature. (3) The γ_1 term in the mean longitude of Enceladus is a forced term due to the eccentricity

* J. Woltjer, *Ann. d. Sternwarte Leiden*, **16**, Part 3, 1928. See also *Bull. Astr. Inst. Neth.*, **1**, 13–16 and 31–36, 1922.

of Dione. The eccentricity of Titan produces a corresponding disturbance of Hyperion.

The differences are quantitative matters and are shown chiefly by the slowness of convergence of series for Hyperion. The disturbances are not so far from linearity that a partly linearized theory similar to that for Enceladus would be seriously in error. Inspection of Woltjer's series shows that higher terms, though not negligible, are much smaller than the main terms. In these conditions it appears that further corrections can be treated by adapting the theory of Enceladus, the chief features having already been treated fully by Woltjer.

Woltjer (p. 66) quotes the following estimates of $1/m_{\mathrm{Ti}} = 1/m'$ found from Hyperion. (a), (b), (c) are his own.

(a)	Mean motion of libration argument	3986
(b)	Mean motion of pericentre	4080
(c)	Mean motion of node	3767 ± 140
(d)	Hill's method (computed by D. Brouwer)	4143
(e)	Eichelberger (special perturbations)	4172 ± 58
(f)	Longitude of pericentre (Samter)	4125
(g)	Libration (Samter)	3910

The uncertainty of (c) is large essentially because the inclination of the orbital planes is small. Eichelberger's result* is from the direct computation of perturbations of the elements, but rests on only about six months' observations. This is only about $\frac{1}{3}$ of the libration period, but as it stands the uncertainty should be valid.

Samter's discussion† is interesting historically because it shares with Eichelberger's the result that a mass $m' = \frac{1}{4700}$ is definitely too low, and also on account of incidental remarks.

Brouwer's method‡ is an up-to-date version of G. W. Hill's. A periodic solution is found by numerical integration over half a synodic period; thus the eccentricity of Titan and the libration are neglected. The motion of the apse and the libration period do not contain the first power of e', and the second power is neglected by Woltjer. On the other hand, the libration, which is large, affects the motion of the apse to an important extent, as pointed out by Samter.

Woltjer assigns weights $\frac{1}{4}$, $\frac{1}{2}$ to (c) and (e). For the others he says that incompleteness of the theoretical computations is the chief source of uncertainty. He assigns weights $\frac{1}{2}$ to (a) and (g), $\frac{1}{3}$ to (b), (d), (f). Adopting these weights gives standard error per unit weight 118 on 6 degrees of freedom, and

$$1/m' = 4033 \pm 71.$$

(The standard error is my computation.) The obvious criticism of this analysis is that the errors of the various estimates are largely due to omissions in computation, and they are unlikely to be independent. It seems to me that Eichelberger's method is the only one not open to this objection, and that its uncertainty cannot be raised to 190 or so without other evidence. The data for (d), (e), (f), (g) are included in those for (a), (b), and the methods of (d), (f), (g) take less full account of higher terms than (a), (b) do. I think, therefore, that all the useful information contained in these solutions is already contained in (a), (b), (c), (e). If only we could be sure that the computations for (a), (b) were complete, they would also include (e).

* W. S. Eichelberger, *Publ. U.S. Naval Obs.* (2), **6**, Appendix 1, B1-17, 1911.

† H. Samter, *Sitz. Preuss. Akad. Wiss.*, 1051-9, 1912.

‡ D. Brouwer, *B.A.N.*, **2**, 119-120, 1924.

Woltjer does not take n and n' as data. Instead, he takes n' and the mean eccentricity of Hyperion. With these and an adopted value of m' the mean motion of pericentre can be calculated, and n is fixed by the consideration that the mean motion of $4l-3l'-\varpi$ must be zero. His expansions are in power of $\mu=m'^{1/2}$, and he starts with the numerical values $m'^{-1}=4033$, $e=0{\cdot}1043$, amplitude of libration 36°. m'^2 and e'^2 are neglected. On these grounds, since e'^2 is about $\frac{1}{1200}$, we might expect that this neglect would not produce errors of more than 6 or so in $1/m'$; but this seems to be far from the case.

Adapting the results for Enceladus, with necessary changes in notation, we should have

$$\nu=\beta n-(3n'-4n), \quad \nu_1=\gamma_1=-\beta' n'+(3n'-4n), \tag{1}$$

$$P=3m'n^2A, \qquad Q=3m'n^2B, \tag{2}$$

$$e=k_0=\frac{m'nA}{\nu}, \qquad k_{10}=0, \tag{3}$$

$$\gamma^2-\nu^2=Pk_0. \tag{4}$$

Woltjer's equation for the mean motion of the libration argument corresponding to (4) is

$$\gamma=1{\cdot}57164\,n'\sqrt{m'}, \tag{5}$$

with no term in m'. The secular terms βn have not been introduced at the start; a correction for them is introduced only at the very end in dealing with the apse motion. In the calculation γ has been supposed expanded in a series $\Sigma\nu_r\mu^r$, and the terms of order μ^2 in the period equation are of the form

$$\mu^2\nu_1^2\frac{d^2\theta_0}{dt^2}=-\frac{\partial^2R_0}{\partial\rho^2}\frac{\partial R_1}{\partial\theta}, \tag{6}$$

where $R_0=M/2a$, θ is the argument $4l-3l'-\varpi$ and R_1 is the part of the disturbing function not containing e'. Hence $\partial R_1/\partial\theta$ contains $m'e$, and the right side of (5) needs a factor $(e/0{\cdot}1043)^{1/2}$ and is analogous to the right of (4). The term ν^2 has been omitted. Now ν is the difference between the actual mean motion of the apse and the part due to the secular terms in the disturbing function, including those due to Titan, and is therefore of order m'. Its neglect is therefore consistent with the neglect of m'^2, but it is not negligible. It can be assessed as follows. Tisserand (vol. 4, p. 108) gives for the effect of Titan

$$\frac{d\varpi}{dt}=m'n\left(4{\cdot}53+\frac{3{\cdot}26}{e}\cos V\right) \tag{7}$$

with $V=\theta$, cos V near -1. The ratio of the terms is $-6{\cdot}91$. The actual mean value is $-18°{\cdot}6562$ in a year, to which the Sun, the ellipticity of Saturn, and the attractions of the other satellites contribute $+0°{\cdot}278$. We have therefore to share the difference $-18°{\cdot}934$ in the ratio 1 to $-6{\cdot}91$, and the separate contributions are $3°{\cdot}20$ and $-22°{\cdot}14$. Thus $\nu=-22°{\cdot}14/1^y$. From observation

$$\gamma=0°{\cdot}562039\pm0°{\cdot}000023/1^d, \tag{8}$$

whence

$$1-\nu^2/\gamma^2=0{\cdot}98836. \tag{9}$$

Then Woltjer's estimate $m'=\frac{1}{3986}$ needs to be revised to

$$m'=\frac{1}{3986}\times0{\cdot}98836(1-\epsilon)=2{\cdot}4796\times10^{-4}(1-\epsilon), \tag{10}$$

where $$e = 0{\cdot}1043(1+\epsilon). \qquad (11)$$

The uncertainty of the observational values is negligible. That of the computations is not, but this correction has probably included the most important effect depending on m'^2.

To adapt the equation for k_0 it might appear that the definition of ν should again take into account the secular term due to Titan, in which case m' should be multiplied by

$$\frac{22^\circ{\cdot}14}{18^\circ{\cdot}93}(1+\epsilon).$$

I think however that this would be erroneous. If other disturbances were absent Woltjer's calculation of the apse motion would be correct; hence the part of βn due to Titan has already been taken into account, and his equation should be right except possibly that the term, already small, due to the other bodies may need a slight change. I therefore write his equation from the apse motion as

$$m' = \frac{1}{4080}(1+\epsilon) = 2{\cdot}4510 \times 10^{-4}(1+\epsilon). \qquad (12)$$

The eccentricity is given as $0{\cdot}10419 \pm 0{\cdot}00025$ (pp. 59, 64), whence

$$\epsilon = -0{\cdot}0011 \pm 0{\cdot}0024. \qquad (13)$$

Direct comparison of the estimates of m' gives $\epsilon = +0{\cdot}0058$, which contradicts the observed value. If we substitute (13) into (10), (12) we get respectively $m' = 2{\cdot}4823 \times 10^{-4}$, $2{\cdot}4483 \times 10^{-4}$. The discordance is about halved, and the mean would be

$$m' = 2{\cdot}4653 \pm 0{\cdot}0170, \quad 1/m' = 4056 \pm 28,$$

on 1 degree of freedom.

The difference is still larger than we should normally expect from the terms neglected. It should be noticed, however, that as the most important part of the perturbations arises near conjunction, the effect of e'^2 is essentially compared with $(a-a')/a$, and may easily be of order 1 per cent. It is likely to be more important in the libration speed than in the apse motion, since the libration requires a further differentiation of the disturbing function. On an average e'^2 will tend to increase negative powers of the distance between the satellites and therefore, probably, to increase the perturbations for a given mass of Titan. On this ground I think that it is incorrect to take the mean of the two estimates, and that allowance for e'^2 might easily change both to about $\frac{1}{4120}$. The uncertainty of this extrapolated value can hardly be put under 1 per cent.

3. *Secular motions.*—Calculated expressions for these are given (logarithmically) by G. Struve (p. 53) in degrees per year. k/a_0^2 and l/a_0^4 are constants depending on the figure of Saturn and the ring. In ordinary form the coefficients of the constants in βn for the various satellites are given in Table III.

TABLE III

	Mimas	Enceladus	Tethys	Dione	Rhea	Titan
k/a_0^2	$1{\cdot}48440 \times 10^4$	$6{\cdot}20243 \times 10^3$	$2{\cdot}93777 \times 10^3$	$1{\cdot}23530 \times 10^3$	$3{\cdot}837 \times 10^2$	$2{\cdot}023 \times 10$
l/a_0^4	$3{\cdot}949 \times 10^3$	$1{\cdot}002 \times 10^3$	$3{\cdot}097 \times 10^2$	$7{\cdot}94 \times 10$	$1{\cdot}265 \times 10$	$1{\cdot}241 \times 10^{-1}$
m_{Mi}	...	$2{\cdot}601 \times 10^5$	$5{\cdot}384 \times 10^4$	$1{\cdot}467 \times 10^4$	...	...
m_{En}	$2{\cdot}947 \times 10^5$	...	$2{\cdot}592 \times 10^5$	$3{\cdot}744 \times 10^4$	$6{\cdot}842 \times 10^3$	...
m_{Te}	$6{\cdot}792 \times 10^4$	$2{\cdot}885 \times 10^5$	...	$1{\cdot}333 \times 10^5$	$1{\cdot}392 \times 10^4$	$4{\cdot}059 \times 10^2$
m_{Di}	$2{\cdot}093 \times 10^4$	$4{\cdot}717 \times 10^4$	$1{\cdot}509 \times 10^5$	...	$4{\cdot}385 \times 10^4$	$7{\cdot}153 \times 10^2$
m_{Rh}	$5{\cdot}849 \times 10^3$	$1{\cdot}019 \times 10^4$	$1{\cdot}862 \times 10^4$	$5{\cdot}183 \times 10^4$	...	$1{\cdot}705 \times 10^3$
m_{Ti}	$3{\cdot}831 \times 10^2$	$5{\cdot}732 \times 10^2$	$8{\cdot}268 \times 10^2$	$1{\cdot}287 \times 10^3$	$2{\cdot}596 \times 10^3$	...
m_{Ja}	...	...	...	...	...	$3{\cdot}162 \times 10^2$

The observed values are given in Table II. For the constants I take the following values.

$$\frac{k}{a_0^2} = 2{\cdot}5 \times 10^{-2}(1 + 10^{-3}\mu_0); \qquad \frac{l}{a_0^4} = 7{\cdot}0 \times 10^{-4}\mu_1;$$

$$m_{\mathrm{Mi}} = 6{\cdot}69 \times 10^{-8}(1 \pm 0{\cdot}030); \qquad m_{\mathrm{En}} = 1{\cdot}46 \times 10^{-7}(1 \pm 0{\cdot}36);$$

$$m_{\mathrm{Te}} = 1{\cdot}141 \times 10^{-6}(1 \pm 0{\cdot}026); \qquad m_{\mathrm{Di}} = 1{\cdot}854 \times 10^{-6}(1 \pm 0{\cdot}033);$$

$$m_{\mathrm{Rh}} = 2{\cdot}5 \times 10^{-6}\mu_2; \qquad m_{\mathrm{Ti}} = 2{\cdot}5 \times 10^{-4}(1 + 10^{-2}\mu_3);$$

$$m_{\mathrm{Ja}} = 2{\cdot}5 \times 10^{-6}\mu_4.$$

In addition we have the equation from the eccentricity of Rhea

$$m_{\mathrm{Ti}} = 2{\cdot}5 \times 10^{-4}(0{\cdot}9716 \pm 0{\cdot}068); \quad \mu_3 = -2{\cdot}8 \pm 6{\cdot}8.$$

In the first place I made a least squares solution from the above data alone. The effects of the first four satellites were calculated from the start; the uncertainties of their masses contribute $\pm 0{\cdot}01$ to the calculated βn for Mimas and Enceladus, $\pm 0{\cdot}017$ for Tethys, and $\pm 0{\cdot}00$ for Dione. Thus these uncertainties can be ignored except for Tethys, and for Tethys they can be treated by raising the standard error of βn from 0·031 to 0·036. The contribution of Iapetus for Titan is $8 \times 10^{-4}\mu_4$. According to the preliminary estimate μ_4 is of order 1, and it would have to reach 4 for its contribution to be as large as the standard error for Titan's apse. At present, therefore, we ignore μ_4.

The first solution gave a rather large χ^2 and made me suspect that some of the smaller uncertainties, especially the very small one for Tethys, might be underestimates. This was checked as follows. The maximum displacements due to equal changes of ϖ and θ would be nearly proportional to ae and $a\sin\gamma$, where a is the radius of the orbit and γ the inclination of the orbit to Saturn's equator; hence $ae\sigma_\varpi$ and $a\sin\gamma\,.\,\sigma_\theta$ should have comparable values in all cases for equally good observing over equal periods. The same should apply to $a\sigma_e$ and $a\sigma_\gamma$, with extensions to the various librations. With a in seconds of arc, the results were mostly between 0″·02 and 0″·03, except for the two for Mimas, which were 0″·05 and 0″·07, and for the node of Titan, which gave 0″·055. The two estimates from Mimas already differ by 2·6 times the standard error of their difference, strongly suggesting that these uncertainties need to be doubled; but none of the standard errors, when so reduced to a common basis of comparison, is ou standingly smaller than the rest. The high accuracy for Tethys is due mainly to the rather large inclination. The large χ^2 was therefore treated by doubling all the smaller observational uncertainties. It was not thought necessary for Tethys to double the contribution of the uncertainties of the masses of other satellites. The equations of condition, so modified, are as follows. I take $\mu_0 = -28 + \mu_0'$, $\mu_1 = 1 + \mu_1'$.

		χ^2
Mimas apse	$0{\cdot}371\mu_0' + 2{\cdot}76\mu_1' + 0{\cdot}01\mu_2 + 0{\cdot}001\mu_3 = +1{\cdot}87 \pm 0{\cdot}20$	0·6
node	$0{\cdot}371\mu_0' + 2{\cdot}76\mu_1' + 0{\cdot}01\mu_2 + 0{\cdot}001\mu_3 = +1{\cdot}50 \pm 0{\cdot}20$	1·2
Tethys node	$0{\cdot}073\mu_0' + 0{\cdot}22\mu_1' + 0{\cdot}05\mu_2 + 0{\cdot}002\mu_3 = +0{\cdot}087 \pm 0{\cdot}065$	0·7
Dione apse	$0{\cdot}031\mu_0' + 0{\cdot}06\mu_1' + 0{\cdot}13\mu_2 + 0{\cdot}003\mu_3 = +0{\cdot}20 \pm 0{\cdot}42$	0·1
Rhea node	$0{\cdot}010\mu_0' + 0{\cdot}01\mu_1' + 0{\cdot}006\mu_3 = +0{\cdot}12 \pm 0{\cdot}08$	1·5
Titan apse	$0{\cdot}000\,5\mu_0' + 0{\cdot}000\,1\mu_1' + 0{\cdot}004\,3\mu_2 = +0{\cdot}007\,8 \pm 0{\cdot}007\,0$	1·0
node	$0{\cdot}000\,5\mu_0' + 0{\cdot}000\,1\mu_1' + 0{\cdot}004\,3\mu_2 = -0{\cdot}001\,4 \pm 0{\cdot}029$	0·0
Rhea eccentricity	$\mu_3 = -2{\cdot}8 \pm 6{\cdot}8$	0·3
		5·4

The normal equations, reduced to unit weight, are as follows.

$$\begin{aligned}
8{\cdot}72\mu_0' &+ 55{\cdot}26\mu_1' + 1{\cdot}32\mu_2 + 0{\cdot}070\mu_3 = + 33{\cdot}5,\\
55{\cdot}26\mu_0' &+ 392{\cdot}4\mu_1' + 4{\cdot}05\mu_2 + 0{\cdot}25\mu_3 = + 236{\cdot}5,\\
1{\cdot}32\mu_0' &+ 4{\cdot}05\mu_1' + 1{\cdot}087\mu_2 + 0{\cdot}026\mu_3 = + 2{\cdot}67,\\
0{\cdot}070\mu_0' &+ 0{\cdot}25\mu_1' + 0{\cdot}026\mu_2 + 0{\cdot}029\mu_3 = + 0{\cdot}180.
\end{aligned}$$

The solution is

$$\mu_0' = +0{\cdot}0 \pm 1{\cdot}9; \quad \mu_1' = +0{\cdot}60 \pm 0{\cdot}28; \quad \mu_2 = +0{\cdot}2 \pm 1{\cdot}5; \quad \mu_3 = +0{\cdot}9 \pm 6{\cdot}1.$$

χ^2 on 4 degrees of freedom is satisfactory, and the data for these satellites are consistent among themselves. The uncertainty of 6 per cent in the mass of Titan, however, calls for improvement, which may be sought in the data for Hyperion and Iapetus.

3.1. *Secular movements of Iapetus.*—The relevant formulae are in H. Struve's paper of 1888. G. Struve gave data up to 1927, but his paper ends abruptly after linear forms have been fitted; presumably he would have brought H. Struve's work up to date had he lived longer. The equations can be written as follows. θ is again written for longitudes on the ecliptic.

$$\begin{aligned}
\sin i \frac{d\theta}{dt} &= -\Sigma K_r \{\sin i \cos i_r - \cos i \sin i_r \cos(\theta - \theta_r)\}\\
&\qquad \times \{\cos i \cos i_r + \sin i \sin i_r \cos(\theta - \theta_r)\},\\
\frac{di}{dt} &= \Sigma K_r \sin i_r \sin(\theta - \theta_r)\{\cos i \cos i_r + \sin i \sin i_r \cos(\theta - \theta_r)\}.
\end{aligned}$$

These were integrated numerically; the values taken for the Sun and for Saturn's equator, referred to the ecliptic and equinox of 1890·0, were

$$\theta_0 = 112^\circ 42', \quad i_0 = 2^\circ 29\tfrac{1}{2}',$$
$$\theta_1 = 167^\circ 58'{\cdot}79 \pm 0'{\cdot}46, \quad i_1 = 28^\circ 04'{\cdot}44 \pm 0'{\cdot}25.$$

The last is based on G. Struve's determinations from the orbital planes of the five inner satellites (Heft 4, p. 48) for 1889·25. These means, weighted according to the stated uncertainties, gave $\chi^2 = 3{\cdot}2$ on 8 degrees of freedom; there was therefore no occasion for modifying the weights, as Struve did. θ_1 was adapted to 1890·0 by allowing for precession. These values could be taken as constant. For Titan the motion of the orbital plane is appreciable. G. Struve's solutions (Heft 5, p. 11) were used. Precession had already been removed from his values of θ_{Ti}. The satellites out to Rhea could be combined with the ellipticity of Saturn, but Titan needed separate treatment. (H. Struve had taken its orbital plane as identical with Saturn's equator, but as these planes can be inclined at about 38′ and the orbit of Iapetus is inclined at about 10° to both, this might have produced an error of order 5 per cent in the mass of Titan.) We have also

$$K_0 = 4'{\cdot}073/1^{\mathrm{y}}, \quad K_1 = 0'{\cdot}694/1^{\mathrm{y}}, \quad K_{\mathrm{Ti}} = 2'{\cdot}779(1 + 0{\cdot}01\mu_3)/1^{\mathrm{y}}.$$

K_1 has a standard error of about 1 per cent, mainly arising from the mass of Rhea, but this can lead to one of only 0·3 per cent in the mass of Titan. The initial values adopted for 1890·0 were $\theta = 142^\circ 05'$, $i = 18^\circ 24'$, and numerical integration was carried out to cover the interval 1770 to 1930. The secular motion is slow and it was enough to regard corrections to the initial values as carried through the entire period. The result was compared with the observed values (adapted to the fixed equinox used). These were much less accurate up to 1880·2 than later.

I took the standard error of θ to be cosec i times that of i in each interval, with the following results.

	Crude residuals		Revised residuals	
	σ_θ	σ_i	σ_θ	σ_i
Up to 1880·2	11′·1	3′·6	11′·0	3′·5
1885–1927	2′·0	0′·62	1′·7	0′·54

The unknowns were estimated with these uncertainties and the standard errors redetermined from the new residuals. No substantial further correction was needed, and the results were

$$\delta\theta(1890{\cdot}0) = +0'{\cdot}56 \pm 0'{\cdot}89,$$

$$\delta i(1890{\cdot}0) = +1'{\cdot}59 \pm 0'{\cdot}35,$$

$$\mu_3 = -5{\cdot}5 \pm 2{\cdot}1.$$

The separate estimates of μ_3 from θ and i were $-4{\cdot}8 \pm 4{\cdot}0$ and $-5{\cdot}7 \pm 2{\cdot}4$. Effectively the uncertainties are on 16 degrees of freedom in comparison with usual standards, since two standard errors have been estimated from 20 equations. The residuals are as follows.

	Observer	O–C	
		θ	i
		′	′
1787·7	Herschel and Bernard	−20·7	−2·7
1832·5	Bessel	−5·7	−4·7
1857·5	Jacob	+9·5	−1·0
1876·7	A. Hall	+1·3	+0·2
1880·2	A. Hall	+7·9	+3·0
1885·6	H. Struve	+0·6	+0·2
1917·2	H. Struve and Bernewitz	+1·7	−0·7
1918·2	H. Struve and Bernewitz	+1·0	+0·4
1926·4	Alden	−3·0	−0·6
1927·4	Alden	+0·1	−0·4

The residuals for the second period are substantially greater than the apparent observational uncertainties would suggest. Since a similar result was found for the inner satellites it looks as if some additional source of disturbance affects secular motions.

It has been indicated above that uncertainty of the mass of Rhea contributes one of about 0·3 per cent to the mass of Titan. The uncertainty of the orbital plane of Titan is about 0′·8, compared with a mutual inclination of 10°, and therefore contributes about 0·1 per cent. The effects of the uncertainty of Saturn's equatorial plane and of changes in the plane of its orbit would be still smaller. Thus the above uncertainty is probably valid and we have from Iapetus

$$1/m_{\mathrm{Ti}} = 4\,243 \pm 93.$$

This value is consistent with Eichelberger's. If, as suggested above, Woltjer's value from Hyperion should be revised to about $\frac{1}{4120}$, it also would be consistent, but a value near $\frac{1}{4033}$ would be hard to reconcile with either the present value or Eichelberger's.

G. Struve gives also data for the eccentricity and pericentre of Iapetus. The pericentre moved 28° in the interval, with reference to a fixed equinox, but

inspection of the residuals left by Struve's linear solution (and indeed the uncertainty that he himself gives) shows that the standard error of $d\varpi/dt$ is not less than 1 part in 30. Since only about a third of the movement is due to Titan it follows that these data cannot give m_{Ti} with an uncertainty under 10 per cent, and detailed computation would be useless. The reason for the lower accuracy is that e for Iapetus is much less than the sines of the inclinations of the orbital planes.

4. *Combination of data.*—The solutions for m_{Ti} are equivalent to the following values of μ_3.

		χ^2
Inner satellites	$+0{\cdot}9 \pm 6{\cdot}1$	0·6
Iapetus	$-5{\cdot}5 \pm 2{\cdot}1$	0·8
Hyperion		
Eichelberger	$-4{\cdot}1 \pm 1{\cdot}3$	0·2
Woltjer (adapted)	$-2{\cdot}9 \pm 1{\cdot}0$	0·4
		2·0 on 3 d.f.

The weighted mean is $-3{\cdot}54 \pm 0{\cdot}74$. The uncertainty given to the adapted value from Woltjer's results is, of course, wholly provisional; but it would be impossible to reject these results and no better course suggests itself. The best method of all, of course, would be to extend the theory of Hyperion to take account of e'^2 and possibly higher terms in m_{Ti}, but it is already clear that this would be an undertaking comparable with the lunar theory.

The above value of μ_3 was substituted in the normal equations derived from the secular movements of the inner satellites. The solution was

$$\mu_0' = +0{\cdot}2 \pm 1{\cdot}9; \quad \mu_1' = +0{\cdot}58 \pm 0{\cdot}28; \quad \mu_2 = +0{\cdot}15 \pm 1{\cdot}5.$$

Comparing with the data that contain these three parameters gives $\chi^2 = 4{\cdot}9$ on 4 degrees of freedom. Then

$$k/a_0^2 = 0{\cdot}024\,305 \pm 0{\cdot}000\,048,$$
$$l/a_0^4 = (11{\cdot}04 \pm 1{\cdot}96)10^{-4},$$
$$m_{Rh} = (0{\cdot}4 \pm 3{\cdot}8)10^{-6},$$
$$m_{Ti} = (2{\cdot}411\,5 \pm 0{\cdot}018\,5)10^{-4},$$
$$1/m_{Ti} = 4\,147 \pm 32.$$

The theoretical values of the secular motions of the inner satellites were finally computed directly from the original formulae with these values. The details are as follows; the separate contributions are given to indicate their relative importance. The third decimal was kept in the calculation in case small positive values might add up.

	Mimas	Enceladus	Tethys	Dione	Rhea	Titan
k	360·78	150·75	71·40	30·02	9·33	0·491 7
l	4·36	1·11	0·34	0·09	0·01	0·000 1
Mi	...	0·02	0·00	0·00		
En	0·04	...	0·04	0·00		
Te	0·08	0·33	...	0·15	0·02	0·000 5
Di	0·04	0·09	0·28	...	0·08	0·001 3
Rh	0·00	0·00	0·01	0·02	...	0·000 7
Ti	0·09	0·14	0·20	0·31	0·63	...
Total	365·40	152·43	72·272	30·60	10·06	0·4943
Obs.	365·60 ± 0·20 365·23 ± 0·20		72·227 ± 0·065	30·75 ± 0·42	10·20 ± 0·08	0·5012 ± 0·007 0 0·492 ± 0·029
O–C	+0·20, −0·17		−0·045	+0·15	+0·14	+·006 9, −0·002

There are some end-figure changes, but recomputation was not thought necessary. G. Struve's values of k and l (p. 54) correspond to

$$k/a_0^2 = 0{\cdot}024\,398; \quad l/a_0^4 = 7{\cdot}013 \times 10^{-4}.$$

These do not appear to have been derived from a least squares solution; the difference in k is significant and that in l is great. He gets a residual of $-0^\circ{\cdot}26/1^y$ for Tethys, in comparison with an apparent "probable error" of $0^\circ{\cdot}021/1^y$, but passes it with no comment except that increasing the mass of Rhea would make it worse. In the present solution this is avoided by having k slightly smaller and l larger, thus fitting both Mimas and Tethys.

In comparison with the discussion of 2.2, the calculated βn for Enceladus gives $\delta(\beta n) = -0{\cdot}34 \pm 0{\cdot}05$. Correcting the masses of Enceladus and Dione accordingly, we have

$$m_{\mathrm{En}} = (1{\cdot}27 \pm 0{\cdot}53)10^{-7},$$

$$m_{\mathrm{Di}} = (1{\cdot}825 \pm 0{\cdot}061)10^{-6}.$$

With regard to the still unsatisfactory determination of the mass of Rhea, this may be interpreted as saying that there is a probability of about $\frac{2}{3}$ that it is less than $\frac{1}{270000}$ of that of Saturn, one of about $\frac{1}{3}$ that it exceeds this value and $\frac{1}{20}$ that it exceeds $\frac{1}{135000}$. As Rhea is brighter than Tethys and Dione a value near $\frac{1}{270000}$ would be plausible.

160 *Huntingdon Road,*
Cambridge :
1953 *January* 9.

THE FIGURES OF ROTATING PLANETS

Sir Harold Jeffreys, F.R.S.

(Received 1953 January 12)

Summary

The theory of the figure of the Earth on the hydrostatic hypothesis, developed by Darwin and de Sitter, is applied to models for Jupiter and Saturn proposed by W. H. Ramsey and B. Miles. It is found that the integral equation for the coefficient of the fourth harmonic is more easily solved directly than by conversion into a differential equation. The results show that the ratio D/J^2, which is equal to $35l/12k^2$ in the treatment of H. and G. Struve, is substantially larger than for a homogeneous body in a hydrostatic state, and in satisfactory agreement with observation for Saturn. Comparison with a model given by Darwin indicates that an inequality stated by de Sitter for the coefficient of the fourth harmonic is not general.

---

1. Darwin, in a classical paper, developed the hydrostatic theory of a rotating planet to the second order in the ellipticity.* His definitions of the quantities concerned led to considerable algebraic complexity; part of this was removed by de Sitter†, but de Sitter published only a few of the formulae and made several errors, some of which have been corrected previously, especially by Bullard.‡ The theory has been applied hitherto only to the Earth, but can now be usefully applied also to Jupiter and Saturn, since the fourth harmonic in their gravitational potentials has an observable effect on the motions of the nearest satellites. Further, the work of Ramsey and Miles on these planets makes it possible to give a theoretical determination of these terms and compare with observation. Application of the first-order theory to rotating stars is perhaps not altogether out of the question.

The first-order terms in the equation satisfied by the ellipticity lead to Clairaut's differential equation, the numerical treatment of which for the Earth is greatly simplified by Radau's approximation. The second-order terms were included by Darwin and de Sitter, who found that they could be treated by a simple modification of the equation. For Jupiter and Saturn, however, Radau's approximation is seriously wrong and a differential equation has still to be solved numerically; and the equation satisfied by the fourth harmonic has so far been solved only by numerical integration, except in a few special cases.

The equations arise in the first place as integral equations, which are converted into differential equations and a pair of boundary conditions. It will appear in what follows that this is unnecessary, and that for the fourth harmonic at least it is easier to apply numerical methods directly to the integral equation.

In the theory of Darwin and de Sitter it was found convenient to define a modified ellipticity, the relation of which to the original ellipticity appears to

* Sir G. H. Darwin, *M.N.*, **60**, 82–124, 1899; *Sci. Papers*, **3**, 78–118.

† W. de Sitter, *B.A.N.*, **2**, 97–108, 1924.

‡ In addition, in (1), 128/105 should be 32/105; in (11), 32/3 should be 8/3. In the set of equations at the foot of the page, $S_1=\frac{5}{3}J$ should read $S_1=\frac{5}{3}J(1+\frac{3}{4}\epsilon)$. (11) is stated to agree with Darwin, but Darwin's form is correct.

contain the coefficient of the fourth harmonic. This surprising result is due to the definition of the ellipticity in terms of the ratio of the equatorial and polar axes. I have proposed for geophysical purposes that the standard ellipsoid should be so defined that, with the outer surface a level surface, it should lead to the correct value of the second harmonic in the external field. Then allowance for an additional fourth harmonic will not affect this relation. The reasons for this course are stronger when the data are purely astronomical.

The following treatment is much shorter than Darwin's; on the other hand, de Sitter's is condensed to an extent that makes the intermediate steps very difficult to reconstruct, and I think a rather fuller account than his is needed.

The axes are taken to be a, a, $a(1-e)$; the sphere of equal volume has radius b, satisfying

$$b^3 = a^3(1-e). \tag{1}$$

The parameter m (de Sitter's ρ_1) is defined by

$$m = \omega^2 b^3/fM. \tag{2}$$

I use the spherical harmonics in the form

$$L_2 = \tfrac{1}{3} - \sin^2\phi' = -\tfrac{2}{3}P_2, \tag{3}$$

$$L_4 = \sin^4\phi' - \tfrac{6}{7}\sin^2\phi' + \tfrac{3}{35} = \tfrac{8}{35}P_4. \tag{4}$$

We have the following identities:

$$\sin^2\phi' = \tfrac{1}{3} - L_2;\ \sin^4\phi' = L_4 - \tfrac{6}{7}L_2 + \tfrac{1}{5};$$

$$\sin^2\phi'\cos^2\phi' = -L_4 - \tfrac{1}{7}L_2 + \tfrac{2}{15};\ L_2^2 = L_4 - \tfrac{4}{21}L_2 + \tfrac{4}{45}. \tag{5}$$

The equation of the ellipsoid, to order e^2, is

$$r/b = 1 - \tfrac{4}{45}e^2 + (e + \tfrac{23}{42}e^2)L_2 + \tfrac{3}{2}e^2L_4. \tag{6}$$

A term $-\kappa\sin^2 2\phi$ would give one in $4\kappa L_4$; taking also, with de Sitter,

$$e' = e - \tfrac{5}{42}e^2, \tag{7}$$

we have

$$r/b = 1 - \tfrac{4}{45}e^2 + (e' + \tfrac{2}{3}e^2)L_2 + (\tfrac{2}{3}e^2 + 4\kappa)L_4, \tag{8}$$

differing from (6) only by a fourth harmonic. The quantity e' is identical with de Sitter's, but, with this definition of e, κ does not appear in the definition of e'.

The internal and external potentials and their first derivatives are continuous at the surface. These relations have to be converted into a pair valid at $r=b$. Suppose that f, $\partial f/\partial r$ vanish at $r=b+h$. Then to the order retained

$$f(r) = \tfrac{1}{2}(r-b-h)^2 f''(b+h) + \tfrac{1}{6}(r-b-h)^3 f'''(b+h), \tag{9}$$

$$\begin{aligned} f'(r) &= (r-b-h)f''(b+h) + \tfrac{1}{2}(r-b-h)^2 f'''(b+h) \\ &= (r-b-h)f''(b) + h(r-b-h)f'''(b) + \tfrac{1}{2}(r-b-h)^2 f'''(b+h), \end{aligned} \tag{10}$$

and to the second order in h, with $r=b$,

$$f(b) = \tfrac{1}{2}h^2 f''(b);\ f'(b) = -hf''(b) - \tfrac{1}{2}h^2 f'''(b). \tag{11}$$

The external and internal potentials of a homogeneous body are taken as

$$\overline{U_0} = fM\left\{\frac{1}{r} + (J+J_0)\frac{b^2}{r^3}L_2 + D_0\frac{b^4}{r^5}L_4\right\} \text{ and} \tag{12}$$

$$U_1 = \frac{fM}{2b^3}(3b^2 - r^2) + fM\left\{(J+J_1)\frac{r^2}{b^3}L_2 + D_1\frac{r^4}{b^5}L_4\right\}. \tag{13}$$

Their difference and its first derivative vanish when r satisfies (8); adapting these conditions to $r=b$ by (11), and treating J as of the first order in e, we find

$$J=\tfrac{3}{5}e';\ J_0=\tfrac{6}{35}e^2;\ J_1=\tfrac{16}{35}e^2;\ D_0=\tfrac{3}{2}e^2+\tfrac{4}{3}\kappa;\ D_1=\tfrac{4}{3}\kappa. \tag{14}$$

Darwin gets equivalent formulae by using the exact potential of an ellipsoid of revolution. The adaptation to a heterogeneous body in a hydrostatic state with external mean radius b is as usual; if r_1 is the mean radius of the surface of equal density through the point considered,

$$\frac{U}{\frac{4}{3}\pi f}=\frac{1}{2}\int_{r_1}^{b}\rho d(3a'^2)+\frac{r_1^3}{r}\rho_0+\frac{3}{5}\left(\frac{Sr_1^5}{r^3}+r^2T\right)L_2+\left(\frac{3}{2}\frac{Pr_1^7}{r^5}+\frac{4}{3}\frac{Qr^4}{r_1^2}\right)L_4, \tag{15}$$

where

$$\left.\begin{aligned}&\rho_0=\frac{1}{r_1^3}\int_{a'=0}^{r_1}\rho da'^3; \qquad S=\frac{1}{r_1^5}\int_{a'=0}^{r_1}\rho d\{a'^5(e'+\tfrac{2}{7}e^2)\};\\ &T=\int_{a'=r_1}^{b}\rho d(e'+\tfrac{16}{21}e^2);\ P=\frac{1}{r_1^7}\int_{a'=0}^{r_1}\rho d\{a'^7(e^2+\tfrac{8}{9}\kappa)\};\\ &Q=r_1^2\int_{a'=r_1}^{b}\rho d\left(\frac{\kappa}{a'^2}\right);\end{aligned}\right\} \tag{16}$$

where in the integrals ρ, e' and κ are regarded as functions of the variable of integration a'. The difference between e' and e does not affect the terms of order e^2, in which the accent is therefore omitted. To the same accuracy the rotation term is

$$\tfrac{1}{2}\omega^2r^2\cos^2\phi'=\tfrac{2}{3}\pi f\bar{\rho}mr^2(\tfrac{2}{3}+L_2). \tag{17}$$

The sum of (16) and (17) must be constant over a surface of constant r'. Then using (8), with b replaced by r_1, and expressing L_2^2 in terms of L_2 and L_4 by (5), we find

$$-(e'+\tfrac{6}{7}e^2)\rho_0+(\tfrac{3}{5}+\tfrac{12}{35}e')S+(\tfrac{3}{5}-\tfrac{8}{35}e')T+\tfrac{1}{2}\bar{\rho}m+\tfrac{10}{21}\bar{\rho}me=0, \tag{18}$$

$$-(\tfrac{1}{2}e^2+4\kappa)\rho_0-\tfrac{9}{5}e'S+\tfrac{6}{5}e'T+\tfrac{3}{2}P+\tfrac{4}{3}Q+\bar{\rho}me=0. \tag{19}$$

Multiplying (18) by $(1-\frac{4}{7}e')$ gives

$$(e'+\tfrac{2}{7}e^2)\rho_0-\tfrac{3}{5}(S+T)-\tfrac{1}{2}\bar{\rho}m=\tfrac{4}{21}e(\bar{\rho}m-3T), \tag{20}$$

and multiplying (18) by $2e'$ and subtracting from (19) gives

$$(3e^2-8\kappa)\rho_0-6eS+3P+\tfrac{8}{3}Q=0. \tag{21}$$

(20) and (21) are de Sitter's equations, except for the correction of an algebraical error. The latter can be checked by comparison with the first-order theory of the fourth harmonic.

2. It is of some interest to show how some of the main results derived from the differential equations can be found more easily directly from the integral equations. I take the first-order equation, for a harmonic of any order, in the form

$$\rho_0\epsilon_n=\frac{3}{2n+1}\left\{\frac{1}{r^{n+3}}\int_{a'=0}^{r}\rho d(a'^{n+3}\epsilon_n)+r^{n-2}\int_{a'=r}^{b}\rho d\left(\frac{\epsilon_n}{a'^{n-2}}\right)\right\}+(0,\ \tfrac{1}{2}\bar{\rho}m). \tag{1}$$

This is essentially an integral equation (written by means of Stieltjes integrals) with fixed termini. For the general case (zero on the right, $n\neq1$) we can without loss of generality take ϵ_n positive at its extreme value; let this be ϵ, taken at $r=r'$ (ϵ_n is continuous by the physical conditions). In (1) replace r by r' and integrate

by parts, which is permissible because ϵ_n is continuous and ρ non-increasing. Then the right side is

$$\frac{3}{2n+1}\left\{r'^{n-2}\frac{\rho(b)\epsilon_n(b)}{b^{n-2}} - \frac{1}{r'^{n+3}}\int_{a'=0}^{r'} a'^{n+3}\epsilon_n\, d\rho - r'^{n-2}\int_{a'=r}^{b}\frac{\epsilon_n}{a'^{n-2}}\, d\rho\right\}, \tag{2}$$

which is not decreased if ϵ_n is replaced by ϵ everywhere (since ρ is non-increasing). Make this substitution and integrate by parts again; then (2) is

$$\leqslant \frac{3\epsilon}{2n+1}\left\{\frac{1}{r'^{n+3}}\int_{a'=0}^{r'}\rho da'^{n+3} + r'^{n-2}\int_{a'=r'}^{b}\rho d\left(\frac{1}{a'^{n-2}}\right)\right\}. \tag{3}$$

The second integral is zero for $n=2$, or $r'=b$, otherwise negative for $n>2$. $\rho_0(r')$ is not a function of a'; hence the right side is

$$\leqslant \frac{3\epsilon}{2n+1}\frac{1}{r'^{n+3}}[\rho_0 r'^{n+3} + \int\{\rho-\rho_0(r')\}\, da'^{n+3}]. \tag{4}$$

Unless ρ is constant there is an r'' such that $\rho(a')\leqslant\rho_0(r')$ for $a'>r''$, $\geqslant\rho_0(r')$ for $a'<r''$; then the integral is

$$\leqslant \frac{n+3}{3}\int\{\rho-\rho_0(r')\}(a'^n - r''^n)\, da'^3 < 0,$$

and finally the right side of (1) is

$$\leqslant \frac{3\epsilon}{2n+1}\rho_0. \tag{5}$$

But this is less than the left side unless $n=1$ or $\epsilon=0$. Hence all harmonics other than those with $n=1$ and the main ellipticity term have zero coefficients.

An argument on these lines is well known to give a proof that $\epsilon_n(b)=0$, after it has been shown from the differential equation that ϵ_n, if not zero, must be monotonic. The above generalization of the argument does not need the preliminary use of the differential equation.

If $n=1$, take $\epsilon_1 = 1/r$. Then the right side of (1) is

$$\frac{1}{r^4}\int_{a'=0}^{r}\rho da'^3 + \frac{1}{r^2}\int_{a'=r}^{b}\rho d(1) = \frac{\rho_0}{r}, \tag{6}$$

agreeing with the left side. Hence $1/r$ is a solution.

The above argument cannot be used directly to prove that it is the only solution, since ϵ_1 may be unbounded. But put $r\epsilon_1=\eta$. Then η is continuous even at $r=0$. Let its maximum be ζ at r'; then the equation (1) is

$$\begin{aligned}\rho_0\eta &= \rho(b)\eta(b) - \frac{1}{r^3}\int_{a'=0}^{r} a'^3\eta\, d\rho - \int_{a'=r}^{b}\eta\, d\rho \\ &\leqslant \left\{\rho(b) - \frac{1}{r^3}\int_{a'=0}^{r} a'^3\, d\rho - \int_{a'=r}^{b} d\rho\right\}\zeta \\ &= \left\{\frac{1}{r^3}\int_{a'=0}^{r}\rho\, da'^3\right\}\zeta = \rho_0\zeta.\end{aligned} \tag{7}$$

Hence $\eta(r')=\zeta$ is false unless the equality sign can be taken in the second line, that is, unless η is a constant. Hence $\epsilon_1 \propto 1/r$ is the only solution (with the trivial exception that in intervals where ρ is constant η might vary without there being any dynamical consequences).

For the main ellipticity term and for the second-order term containing a fourth harmonic, the terms of lowest order take the form

$$\rho_0\epsilon_n = \frac{3}{2n+1}\left\{r^{n-2}\frac{\rho(b)\epsilon_n(b)}{a^{n-2}} - \frac{1}{r^{n+3}}\int_{a'=0}^{r} a'^{n+3}\epsilon_n d\rho' - r^{n-2}\int_{a'=r}^{b}\frac{\epsilon_n}{a'^{n-2}}d\rho'\right\} + f(r). \tag{8}$$

Take as a trial solution

$$\epsilon_n = \eta_n = \frac{f(r)}{\rho_0} \tag{9}$$

and substitute on the right; then the modulus of the first set of terms, by the previous argument, is

$$\leqslant \frac{3}{2n+1}\rho_0|\eta_n|_{\max}; \tag{10}$$

hence the correction is everywhere numerically $\leqslant \frac{3}{2n+1}(\eta_n)_{\max}$ and the results of successive approximations converge like $\sum_m\left(\frac{3}{2n+1}\right)^m$. Since the trial solution is continuous the corrections are continuous, and the limit, being the sum of a uniformly convergent series of continuous functions, is continuous; hence all the required conditions are satisfied.

I have not succeeded in finding a direct proof from the integral equation that ϵ_2 is monotonic for the main ellipticity term. Clearly ϵ_2 is everywhere >0, since the first approximation and all corrections are positive.

The series may converge too slowly for direct computation. In fact if the upper bound of the first approximation is ϵ, that of the true solution may be as large as $\frac{2n+1}{2(n-1)}\epsilon$, which is $\frac{5}{2}\epsilon$ for $n=2$ and $\frac{7}{4}\epsilon$ for $n=3$. The former is actually attained for a homogeneous body (for which the first approximation is $\frac{1}{2}m$). This difficulty may be circumvented as follows. Write the equation as

$$\epsilon_n = \frac{f(r)}{\rho_0} + K\epsilon_n, \tag{11}$$

where K is a linear operator. The first approximation is $\epsilon_n = u_1 = f(r)/\rho_0$. Substitution of u_1 on the right gives a second approximation u_2. The essence of the matter is that $\left\{u_m - \frac{f(r)}{\rho_0} - Ku_m\right\}$ forms a sequence of functions, decreasing approximately geometrically. With two values, say $m=1, 2$, we can choose a constant μ so that the function

$$\left|(1+\mu)\left(u_2 - \frac{f(r)}{\rho_0} - Ku_2\right) - \mu\left(u_1 - \frac{f(r)}{\rho_0} - Ku_1\right)\right| \tag{12}$$

is everywhere as small as possible, and this will be done by taking μ so that the greatest positive and negative values of the expression under the modulus sign are equal in modulus. But this amounts to saying that, judged by the residuals left in the original equation,

$$(1+\mu)u_2 - \mu u_1 \doteqdot \frac{f(r)}{\rho_0} + K\{(1+\mu)u_2 - \mu u_1\} \tag{13}$$

with the smallest possible error. Then $(1+\mu)u_2 - \mu u_1$ can be used as the first of a second pair of approximations, and solution should be rapid. The method

is essentially one of relaxation. In contrast to the use of such methods for differential equations, the "point adjustment" needed to smooth out rapid fluctuations of the errors does not arise, being automatically taken into account in the integration process; all the approximations are "block adjustments". On the other hand, it may not be so convenient for most integral equations as with the present ones, because these have the feature that the kernel can be written in the form

$$g_1(a')h_1(r)+g_2(a')h_2(r),$$

thus containing four functions that can be tabulated once for all; a more general type of kernel would have to be fully tabulated as a function of two variables.

3. *Possible variation of* D/J^2.—The form of a homogeneous rotating body would be an exact ellipsoid of revolution. It was shown by H. Struve that his parameters k and l would lead to

$$l/k^2=\tfrac{10}{7}$$

for all ellipticities, corresponding, in the present notation, to $D/J^2=\frac{25}{6}$. The value found in my discussion of Saturn's satellites was $l/k^2=1{\cdot}87\pm0{\cdot}33$, and it is desirable to examine whether the doubtfully significant difference would be explicable by variation of density. At first sight this seemed unlikely. For if we consider an extreme case where the whole mass is concentrated in a homogeneous core, the outer part has no dynamical effect at all, the core would take a Maclaurin form, and the same relation would hold. Thus rather wide departures from homogeneity might not affect D/J^2. Examination of Bullard's solution for the Earth, however, did show an appreciable departure, and direct examination of the data for the constitutions of Jupiter and Saturn was desirable.

3.1. *Solution for Jupiter.*—Possible distributions of density in Jupiter and Saturn have been given by Ramsey and Miles.* They worked out the second-order theory of the ellipticity, but not the fourth harmonic. The details were not published, but Ramsey and Miles kindly gave me a typed copy. Of their models, those denoted by J_3 and S_3 agreed best with the observational data; but those denoted by J_4, S_4 were only slightly inferior and have the advantage for the present purpose that the distributions are smoother. The object of the present work is to find out whether the ratio D/J^2, for given m, can depart widely from the value $\frac{25}{6}$ that it takes for a homogeneous body of any ellipticity. Accordingly I accepted their values of η derived by de Sitter's method for J_4, S_4, derived the distribution of e, and then used the integral equation 2 (21) to estimate κ. I use their values for the mass, radius, and rate of rotation of each planet. The results for Jupiter are in Table I, where $m=0{\cdot}083\,39$.

κ at most is only about $0{\cdot}1\,e^2$; this is mainly because the e^2 terms in the integral equation are (in much of the interval) nearly as 1, -2, 1. The corrections in the second approximation never exceeded 0·1, so that the method is even more successful than there was apparent reason to expect. This is chiefly because κ is small except where ρ is small compared with $\bar{\rho}$.

For the external field

$$\bar{\rho}b^5J=\frac{3}{5}\int_{r=0}^{b}\rho\,d\{r^5(e'+\tfrac{2}{7}e^2)\},$$

$$\bar{\rho}b^7D=\frac{3}{2}\int_{r=0}^{b}\rho\,d\{r^7(e^2+\tfrac{8}{9}\kappa)\}.$$

* W. H. Ramsey, *M.N.*, **111**, 427–447, 1951. B. Miles and W. H. Ramsey, *M.N.*, **112**, 234–243, 1952.

For comparison with a homogeneous body it seemed best to drop the e^2 term in J. This is because e^3 and κe are neglected in D, and we know that for a homogeneous body D/J^2 is independent of e. It therefore appeared possible that inclusion of a higher term in J and not in D would actually decrease the accuracy. On this basis $J=0{\cdot}0260$, $D=0{\cdot}00339$, of which only $\frac{1}{20}$ comes from κ. Then

$$D/J^2=5{\cdot}0.$$

Allowance for the e^2 term would reduce this to 4·8 or 4·9.

TABLE I

$r(10^9$ cm)	η	e	ρ	ρ_0	$10^4\kappa$ (first approximation)
6·99	1·099	0·0676	0·11	1·33	3·2
6·96	1·065	0·0673	0·14	1·35	3·1
6·9	1·034	0·0667	0·22	1·38	2·9
6·8	0·977	0·0657	0·27	1·43	2·6
6·7	0·926	0·0648	0·30	1·48	2·4
6·5	0·801	0·0632	0·36	1·60	2·0
6·3	0·691	0·0617	0·39	1·71	1·5
6·10	0·565	0·0605	0·42	1·84	1·3
6·10	0·565	0·0605	0·89	1·84	1·3
6·0	0·533	0·0599	0·96	1·89	1·2
5·5	0·416	0·0575	1·28	2·12	1·0
5·0	0·324	0·0555	1·60	2·34	0·7
4·5	0·254	0·0539	1·92	2·56	0·6
4·0	0·194	0·0524	2·23	2·77	0·5
3·5	0·143	0·0513	2·53	2·97	0·4
3·0	0·118	0·0503	2·80	3·14	0·3
2·5	0·105	0·0492	3·05	3·30	0·2
2·0	0·047	0·0484	3·27	3·42	0·1
1·5	0·025	0·0479	3·43	3·52	0·1
1·0	0·020	0·0475	3·56	3·59	0·0
0·0	0·000	0·0472	3·66	3·66	0·0

TABLE II

r	η	e	ρ	ρ_0	$10^4\kappa$
5·755	1·038	0·1122	0·125	0·713	9·8
5·6	0·968	0·1091	0·25	0·756	8·4
5·4	0·902	0·1055	0·31	0·810	7·1
5·2	0·841	0·1021	0·35	0·867	5·6
5·0	0·774	0·0989	0·39	0·929	4·7
4·5	0·542	0·0923	0·45	1·117	2·2
4·08	0·247	0·0888	0·49	1·337	1·1
4·08	0·247	0·0888	1·00	1·337	1·1
4·0	0·238	0·0884	1·03	1·355	0·7
3·5	0·183	0·0859	1·21	1·469	0·6
3·0	0·123	0·0839	1·38	1·585	0·6
2·5	0·092	0·0823	1·53	1·676	0·5
2·0	0·061	0·0809	1·66	1·763	0·3
1·5	0·038	0·0797	1·77	1·834	0·2
1·0	0·015	0·0791	1·85	1·87	0·0
0·0	0·000	0·0785	1·91	1·91	0·0

3.2. *Solution for Saturn.*—This is given, for Model S_4, $m=0{\cdot}1353$, in Table II. Further corrections were negligible. The results lead to

$$J=0{\cdot}0438;\ D=0{\cdot}0116;\ D/J^2=6{\cdot}0.$$

The conjecture that D/J^2 for a body in hydrostatic state is nearly independent of the distribution of density is therefore seriously wrong for Jupiter and Saturn.

4. *A model given by Darwin.*—To investigate the possible range of values of D/J^2 I consider a case treated by Darwin, of a body of unit density with a fraction μ of its mass concentrated at the centre, the density otherwise being uniform. I take the radius also as 1. Darwin determined the value of μ that makes κ greatest, m being taken as given, but did not explicitly attend to the external field. (My μ is Darwin's $\mu/(1+\mu)$.) The results above become

$$\rho_0 = \frac{1}{r^3}\{\mu + (1-\mu)r^3\},$$

$$S = (1-\mu)r^2(e' + \tfrac{2}{7}e^2); \quad T = (1-\mu)(e_s' + \tfrac{16}{21}e_s^2 - e - \tfrac{16}{21}e^2),$$

$$P = (1-\mu)(e^2 + \tfrac{8}{9}\kappa); \quad Q = (1-\mu)(\kappa_s r^2 - \kappa).$$

To the first order

$$e_s' = \frac{\frac{5}{4}m}{1+\frac{3}{2}\mu}, \quad e' = e_s' \frac{r^3}{\mu + \frac{8}{5}(1-\mu)r^3 - \frac{3}{5}(1-\mu)r^5},$$

and to the second order

$$\kappa_s = \frac{9\mu e_s^2}{8(2+\mu)},$$

$$J = \tfrac{3}{5}S, \; D = \tfrac{3}{2}P,$$

$$\frac{D}{J^2} = \frac{25}{6}\frac{2+2\mu}{(1-\mu)(2+\mu)}.$$

Hence this ratio actually tends to infinity as μ tends to 1, and is never less than $\frac{25}{6}$, the value for a homogeneous body. The reason is that as $\mu \to 1$, J is of order $(1-\mu)m$ and D of order $(1-\mu)m^2$.

This result is of some interest in relation to the models J_3 and S_3 of Miles and Ramsey, which they thought the most probable. In these the heavy elements are supposed concentrated near the centre. For J_3 the nucleus has about $\frac{1}{20}$ of the mass of the whole body, and we might expect D/J^2 to be increased by about 7 per cent above the value found for J_4. For Saturn the corresponding quantities are about $\frac{1}{14}$ and 10 per cent. It is therefore quite likely that the correct values are about 5·4 for Jupiter and 6·6 for Saturn.

5. *Bullard's solution for the Earth.**—This applies de Sitter's method. Since the differential equation for κ is derived from Darwin's solution, it is correct. The boundary condition (10) for κ should read

$$\frac{d\kappa}{d\beta} = -4\kappa - \tfrac{5}{4}\rho_1'\epsilon_1' + \tfrac{25}{16}\rho_1^2;$$

the calculations have been carried out with the correct form. The easiest way of finding D/J^2 is to compare with the International Formula, which gives $J = 0{\cdot}00164$, $D = 0{\cdot}0000107$, $D/J^2 = 3{\cdot}98$. Bullard finds on the hydrostatic theory that $\kappa = 6{\cdot}8 \times 10^{-7}$. This represents the distortion of the level surface, and since the International Formula is consistent with the outer surface being an exact ellipsoid, irrespective of whether hydrostatic conditions subsist in the interior, 4κ is added directly to D and makes $D/J^2 = 4{\cdot}98$.

6. *An inequality given by de Sitter.*—De Sitter states without proof that for surface values

$$\kappa \leqslant \tfrac{5}{16}me - \tfrac{1}{4}e^2.$$

* E. C. Bullard, *M.N., Geophys. Suppl.*, **5**, 186–192, 1946.

This relation is satisfied by Bullard's solution for the Earth and by the above solutions for J_4 and S_4. For Darwin's model, on the other hand, the right side is $\frac{3}{8}\mu e^2$, which is less than κ_s except for $\mu=0$ and $\mu=1$. It appears, therefore, that the above inequality is not general.

7. *Comparison with observation for Saturn.*—The values for k and l for Saturn* made $k/l^2=1{\cdot}87\pm0{\cdot}33$; to reduce to J and D we must multiply by $\frac{35}{12}$, giving $D/J^2=5{\cdot}45\pm0{\cdot}96$. Thus the theoretical results for Saturn are consistent with observation. There is no basis for an attempt to estimate the mass of the Ring, such as might have been made had the observational value exceeded the theoretical one.

On the other hand, the results of the present paper suggest that it is reasonable to take for Saturn $D/J^2=6{\cdot}3\pm0{\cdot}3$, which is equivalent to $l/k^2=2{\cdot}16\pm0{\cdot}10$. This leads to an additional equation of condition, in the notation of the previous paper,

$$0{\cdot}027\mu_0'-7{\cdot}0\mu_1'=-5{\cdot}7\pm0{\cdot}58.$$

Combining this with the previous data I get

$$k/a_0^2=0{\cdot}024\,270\pm0{\cdot}000\,016;\quad l/a_0^4=(12{\cdot}42\pm0{\cdot}52)10^{-4};$$

$$m_{Rh}=(2{\cdot}8\pm2{\cdot}8)10^{-6};\quad m_{Tl}=(2{\cdot}412\pm0{\cdot}018)10^{-4}.$$

The comparison with observation is as follows.

	Observation		Calc.	χ^2
Mimas apse	365·60	±0·20	365·44	0·6
node	365·23	±0·20	365·44	1·1
Enceladus			152·37	...
Tethys node	72·227	±0·065	72·257	0·2
Dione apse	30·75	±0·42	30·69	0·0
Rhea node	10·20	±0·08	10·05	3·5
Titan apse	0·5012	±0·0070	0·4976	0·3
node	0·492	±0·029	0·4976	0·0
μ_3 from Rhea	−2·8	±6·8	−3·5	0·0
$1{\cdot}85\mu_3$ from Hyperion and Iapetus	−6·6	±1·4	−6·5	0·0
l/k^2 (theory)	2·16	±0·10	2·11	0·2
				5·9 (6 d.f.)

The general comparison is quite satisfactory. In so far as the theoretical value taken for l/k^2 is the mean of values for the models S_3 and S_4, and the other data indicate a most probable value a shade less than this, it may be held that the result slightly favours S_4, but the evidence is weak.

St. John's College,
Cambridge:
1953 *January* 9.

* H. Jeffreys, *M.N.*, **113**, 81–96, 1953.

GLACIAL EUSTASY AND THE ROTATION OF THE EARTH*

Andrew Young

(Received 1952 December 29)

Summary

A rise in sea-level at the rate of 1 cm per century following a reduction in the extent of ice in glaciated regions results in an increase of the moment of inertia of the Earth about its axis of rotation sufficient to lengthen the day by 10^{-4} second per century provided there is no isostatic compensation. The consequent apparent secular acceleration of the Moon is 2 seconds of arc per century per century. The changes however are negligible if immediate and complete isostatic compensation occurs. A tentative discussion of climatic and glaciological evidence indicates that fluctuations of sea-level do occur and that their effect on the Moon's apparent acceleration may be appreciable. This re-opens the question of the cause of the Moon's apparent acceleration, which is usually attributed to the deceleration of the Earth by tidal friction. The change in tidal friction due to a rise in sea-level is very small and has not been sufficient to produce the observed change in the Moon's acceleration.

* The full text of this paper is published in *M.N., Geophys. Suppl.*, **6**, No. 7, 1953.

OBSERVATIONS OF COLOUR TEMPERATURES OF STARS*

Vol. II. Relative Gradients

Royal Observatory, Greenwich

(Communicated by the Astronomer Royal)

(Received 1953 March 13)

Summary

Relative spectrophotometric gradients are given for 250 stars, comprising practically all stars of HD types Oe5, B and A, brighter than $4^m{\cdot}5$ and N. of the equator, as well as a selection of F- and G-type stars. The observations, which were made photographically with slitless spectrographs attached to the 30-in. and 36-in. reflectors at Greenwich, consist essentially in comparing (at the same altitude) the blue/red intensity ratio in the spectrum of a programme star with the corresponding ratio in one or more of 25 standard stars well distributed round the northern sky. The plates were photometrically standardized in an auxiliary spectrograph using multiple slits calibrated by the method of full and half apertures.

The intensity measures in the stellar continuum were made at two wavelength regions: one at 6321 A (harmonic mean), the other either at 4285 A (for O and B stars) or 4270 A (for A stars) or 4693 A (for later types). Corrections were made to the measured gradients to allow for residual atmospheric reddening (the comparisons being made, however, as far as possible at equal altitudes) and for inequality in exposure time between star spectrum and calibration spectrum.

The zero of gradient is defined arbitrarily as the mean gradient of nine specified stars of type A0. An estimate of the colour temperature corresponding to this zero is given from comparisons between the stars and a standard lamp operating at known temperature. This enables any listed gradient to be converted to a temperature if desired, but temperature values are not tabulated because of the uncertainty of the terrestrial comparisons.

Detailed comparisons are made with other measures of star colour, including earlier gradient determinations, colour equivalents, photoelectric colours and photographic colour indices. Absorption lines necessarily comprised in the broad wave-length ranges selected by colour filters markedly affect the correlation with photographic and photoelectric colours. The relation between gradient and spectral type is investigated using both HD and Mt. Wilson types; and the cause of the abnormally red colour of certain early-type stars is discussed. A special series of observations of δ Cephei enables a comparison to be made between the light curve and a curve showing the variation of gradient with phase.

* The full text of this publication appears as a separate volume of the same title published by H.M. Stationery Office, London, 1952; price £1 5s. net.

CONTENTS

PAGE

Meeting of 1953 January 9:
Fellows elected ... 1
Junior Members elected ... 2
Presents announced ... 2
Award of the Gold Medal to Professor S. Chandrasekhar ... 2
Award of the Eddington Medal to Canon Georges Lemaître ... 2
Award of the Jackson-Gwilt Medal and Gift to Mr John Philip Manning Prentice ... 2

P. A. G. Scheuer and M. Ryle, An investigation of the H II regions by a radio method ... 3

R. d'E. Atkinson, Cinematography of partial solar eclipses. I. General ... 18

D. W. Sciama, On the origin of inertia ... 34

D. C. Pack, The motion of a gas cloud expanding into a vacuum ... 43

Ida W. Busbridge, Coherent and non-coherent scattering in the theory of line formation ... 52

Humphry M. Smith, Quartz clocks of the Greenwich Time Service ... 67

Sir Harold Jeffreys, On the masses of Saturn's satellites ... 81

Sir Harold Jeffreys, The figures of rotating planets ... 97

Summary of paper published in the *Geophysical Supplement:*
Andrew Young, Glacial eustasy and the rotation of the Earth ... 106

Summary of paper published by H.M. Stationery Office:
Royal Observatory, Greenwich, Observations of colour temperatures of stars. Vol. II, Relative gradients ... 107

Printed in London by Taylor & Francis, Ltd., 1953 June.

MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 113 No. 2 1953